UNIVERSITY OF CALIFORNIA, SAN DIEGO

Identify Inhibitors on DAPK2 and ELMO1-DOCK2 Complex via Virtual Screening

A thesis submitted in partial satisfaction of the requirements for the degree

Master of Science

in

Chemistry

by

Meixin Wu

Committee in charge:

Professor Wei Wang, Chair
Professor Rommie Amaro
Professor Michael Gilson

2016

The Thesis of Meixin Wu is approved and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2016

Dedication

To the colleagues in Dr.Wang's lab whom I appreciate for their supports and collaborations. Especially to Dr. Wei Wang and Dr. Nan Li whose guidance and wise-advices shaped who I am now as a graduate student.

To my family for unconditional love and support.

Table of Contents

List of Figures

List of Tables

ABSTRACT OF THE THESIS

Identification Inhibitors on DAPK2 and ELMO1-DOCK2 Complex via Virtual Screening

by

Meixin Wu

Master of Science in Chemistry

University of California, San Diego 2016

Professor Wei Wang, Chair

Death associated protein kinase 2 (DAPK2) is a serine threonine kinase with 370 amino acid residues, whose structure is highly conserved with the other members in the death associated protein kinase (DAPK) family. To find out the inhibitors with high specificity to DAPK2, structural based virtual screening was carried out. According to the results of docking, ligand clustering, MM/GBSA analysis and position of ligands in the DAPK2 binding pocket, it can be concluded that DA7, whose binding affinity and MM/GBSA score are -11.7 Kcal/mol and -47.5 Kcal/mol respectively, is the most ideal candidate so far for DAPK2.

Engulfment and cell motility 1 protein (ELMO1) is a human protein regulated by

ELMO1 gene, with 720 residues in total, which is critical in clearing the apoptotic germ cells in vivo. Dedicator of cytokinesis 2 (DOCK2) is a guanine nucleotide exchange factor with a mass of 213 kDa, which is responsible for the activation of the chemotaxis process. DOCK2 and ELMO1 form a large, rigid complex via the SH3 domain at DOCK2 or the formation of the five-bundle helices by ELMO1 and DOCK2. To identify the potent inhibitors that interrupt the formation of the ELMO1-DOCK2 complex, both structural- and ligand- based virtual screening were carried out at the two potential pockets: one is located at the SH3 domain of DOCK2 and the other is located between the 5-helix bundles. From the docking results, ZINC22013692 and ZINC48368120 were candidates with the highest binding affinity for the pocket at the SH3 domain and 5-helix bundles respectively.

Chapter 1 Introduction

I. Virtual Screening

1. General Background

Virtual screening provides a powerful tool to identify potential drug candidates with manageable size in a low cost and time-efficient way. [1][2] Depending on the level of receptor information used in the calculation, virtual screening methods can be divided into ligand-based and structure-based methods. [3][4] Ligand-based methods focus on the usage of small molecules alignment between a large database and the reference structure, while structure-based approach imitates the binding activity between ligand and the target receptor based on their 3D information. [4]

2. Ligand-Based Method

Ligand-based virtual screening is a ligand search method aiming at finding molecules highly similar to the known receptor-binding ligands by assuming that similar compounds could have similar or better binding affinity to the receptor. [4] A database search is usually performed to find potential compounds using certain pre-defined similarity metric. These metric usually consider the similarity between the structure of the two molecules, such as the two dimensional sub-structural similarity or the three-dimensional pharmacophore model. [4][5]

There are at least three types of ligand-based screening methods: small molecule alignment, pharmacophore-based and descriptor-based. These methods use different

algorithms such as structure, individual fragments or electrical properties of the reference molecule to generate ligand candidates, and then the candidates were ranked based on how similar they are to the reference molecule.

Several methods use the sub-structure alignment to measure the similarity between molecules [6]- [10], such as Flexs, GASP (Genetic Algorithm Similarity Program), MEP (Molecular Electrostatic Potential), MIMIC, and fFlash. The primary difference between these methods is whether the flexibility is considered on the ligands. In Flexs, the conformation of query molecule is fixed and the ligands in the database are in flexible state. GASP allows flexibility on both query and database ligands. MEP uses both sub-structure alignment and molecular field comparison to evaluate the similarity between ligands. The flexibility is considered on both query and database ligands by the genetic algorithm to get the optimal alignment. Only after getting the optimal alignment, the molecular electrostatics fields of the two molecules are compared for the similarity. Both MIMIC and fFlash use the molecular field concept similar to MEP, but these two methods only allow the database ligands to undergo any conformational change.

Another group of methods use various descriptors to characterize the molecular properties of the ligands. The simplest descriptor contains the one dimension information for the ligand, such as molecular weight, molar refractivity, and log P [4] [11]. More sophisticated descriptors use the vector of binary or continuous values, to describe the molecular property. Among such methods, MACCS keys [13]- [14], Pubchem fingerprints [15], and sub-structure fingerprint [12] use Boolean value to describe the substructure of the molecule; Daylight [16] and UNITY [12] use hash-based fingerprints to describe the atomic connection of the molecule; Volsurf [17] evaluate the molecular field of the molecule

using real values; MTree [18] and NIPALSTREE [19] use the more complicated tree-type data structure to better describe the molecule.

Another worth-mentioned method is the pharmacophoric model. Programs like ALMOND [20] provide information such as the internal geometric relationship between the ligand and the receptor and the pharmaco-dynamic properties of the ligand-receptor complex. [4] Other pharmacophoric software are designed based on different principles, which can be the properties of the location for the specific substructure of the molecule, point comparison, and rigid body molecular superposition. [21]- [25]

3. Structure-Based Method

In structural-based approach, there are generally four steps for the ligand candidate search: the identification of the binding pocket on the target receptor, the preparation of the ligand database, running the virtual screening process using docking software, and rank the docking results by scoring function. Unlike ligand-based screening, which relies on the similarity of the ligand candidates to the referenced molecule, the ligand candidates was docked into the target receptor and the ligand was ranked based on the binding affinity of the candidates to the receptor. Docking is a process that combines the receptor and the ligand into a complex at the target site of the receptor. [1][3][4] In general, the docking method is consisting of two steps: searching and scoring.

3a. Searching Methods

The searching of the conformational space usually uses stochastic algorithms like Monte Carlo, simulated annealing and genetic algorithm are used to optimize the small molecule conformation in the pre-defined binding pocket. [4] [12] [26] In Monte Carlo

algorithm, the conformation of each ligand is assessed based on Boltzmann probability. [27] [28] [29] At a given temperature, a certain amount of random conformations were produced and arranged in a Markov chain, meaning that the evaluation of the current conformation only depends on its previous one. [1] [28] [29] To determine which confirmation should be retained, Monte Carlo algorithm uses Metropolis criterion: If the new conformation have a better score than the previous one, then it is accepted; otherwise the acceptance probability is related to the "temperature" of the system. [1][4] [28] [29] Another widely used algorithm in conformational search is genetic algorithm, adapted from the theory of evolution that mimics the process of biological competition. [1][4] [30] In the genetic algorithm, the parameters are packed stochastically in a "chromosome", which produce solutions to the research problem and are assessed by the fitness function. [1] [32] "Chromosomes" that yield the best solutions will be crossover and mutated similar to genetic recombination and mutation to produce the next generation. [1][4] In terms of docking, the values representing the mass center coordinate, the relative orientations to the receptor, and the sub-structures of the ligand are defined as the ligand's state variable and packed into the "chromosome". [1] [4] During the searching, chromosomes carrying the state variables of the top ligand candidates are crossovered and mutated. [1] The binding conformation with the best score throughout the GA search is output as the final docking result. The docking applications using the genetic algorithm include DOCK, Autodock and Darwin [30].

Lamarckian genetic algorithm is a modified genetic algorithm designed based on the principle of biological evolution and Mendelian genetics. [31] In LGA, the genotype of the parents is initially transformed into its phenotype via a mapping function, followed by

the utilization of a local search of the parental phenotype and the transformation of the parental phenotype into the genotype of the next generations via a Lamarckian inverse mapping if the local minima hits. [31] In terms of docking, binding energy calculated from the coordinates of the ligands is treated as the phenotype and the ligand's state variable described in genetic algorithms are treated as genotype. [31] AutoDOCK is the docking program combining both GA and LGA to perform the conformational search of the ligand candidates. [4] [31]

An alternative way of conformational search is the incremental construction search, which breaks the ligands at their rotational bonds and docks the fragments of these ligands rigidly at each favorable position of the binding site, with the largest fragment selected as the base one. [4] The orientations of the base fragment docked at the binding pocket is retained, followed by the addition of other smaller fragments in different orientation and the assessment via the designated scoring function. [4] This process is repeated until the entire ligand is assembled. [4] Docking software developed by incremental construction are FlexX [32], Hammerhead [33] and HOOK [4].

3b. Scoring Functions

After the selection of the ligand candidates, a score function is used to differentiate the poses of the ligand candidates at the binding site and rank the candidates in accordance to their calculated scores. All scoring function design is based on the hypothesis that the native complex conformation has the global minimum score. Scoring functions can be classified into four categories: force field based, empirical-based, knowledge-based, and consensus-based. [1][4]

Force-field based scoring functions estimate the sum of the contributions from interactions with clear physical basis, such as hydrogen bond, electrostatic interaction and Van der Waals interaction between ligand and the target receptor. [1][4] When scoring the relative binding strength of the ligand, only the enthalpic gas phase for energetics and structures are considered. [1][4] Therefore to explain the desolvation effect caused by the water molecule in virtual screening, PB/SA and GB/SA analysis are applied to calculate the electrostatic desolvation free energy and the free energy caused by hydrophobic interaction respectively. [34] DOCK is a typical docking software; whose scoring function is AMBER force-field based. [34]

Empirical scoring functions estimate the binding energy from various terms, whose coefficients can be obtained from regression analysis such as experimental data and X-ray crystallography. [1][4] Although method of the binding energy estimation is similar to the force-field scoring function, empirical scoring function can include the non-enthalpic terms. [1][4] One of the disadvantages of empirical based scoring function is that since it depends on the experimental data set used to perform linear regression analysis, the terms are not evenly weight. [4]

Nowadays there are several software developed based on the principle of Empirical-based scoring function, which is adapted from the construction of statistical preference of the atom pairs in terms of the Helmholtz energy and Boltzmann constants. [4] Glide, RosettaScore, LUDI, ChemScore, Score, X-Score, and F-Score are software that falls into these categories. [4] [34]

Knowledge-based scoring functions estimate binding score via the reproduce experimental structures using potentials from the interaction of atom pairs. [1][4] In

knowledge-based scoring function, the protein-ligand complex is modeled via the atomic interaction potential in pairwise. This scoring function is efficient in screening large ligand database since the computational method is relatively simple, and therefore not sensitive to the training data set. [1][4] Compared with the force-field based and empirical-based scoring functions, it can balance the relationship between the computational speed and accuracy. [34] The potentials calculated from the inverse Boltzmann relationship are based on the structures instead of the reproduction of the binding affinities by fitting. [34] In addition, the scoring progress of the atomic potential in pairwise fashion can be fastening to a level similar to the empirical based ones. [4] [34] Docking software designed based on the knowledge-based scoring functions are SMoG, PMG,ITScore, DrugScore and BLEEP. [4] [34]

Additionally, consensus scoring technique can be applied by combining the existing scoring functions to minimize the errors and increase the accuracy of finding optimal candidates. [1][4] In consensus scoring function, the construction of the consensus score strategy from each individual score is crucial for identifying the true candidates. [34] Applications using consensus scoring function are Multi-Score, CSC, and SeleX-CS. [1] [34]

4. AutoDock Vina

4a. Origin of AutoDock Vina

As the updated version from Autodcok 4, AutoDOCK Vina is able to perform different levels of flexible docking through assigning freedom of rotation on different types of rotatable bonds on ligands. Autodock Vina implements a completely different set of empirical scoring functions from the force field like scoring functions used in Autodock4. [35] In Autodock, as described section **3.1.1a.** and **3.2.**, it uses stochastic

algorithms such as Monte Carlo simulated annealing, GA and LGA to perform the conformational search of the ligand candidates. [31] However in Vina it uses a limited Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm to perform the conformational search for ligand candidates. [36] In Autodock 4 a stepwise fitness function is used to perform the local search of the ligand conformation while in Vina a gradient is applied in the scoring function. According to Chang et.al, Autodock4 is less consistent in the search of the energy minima than Vina when evaluating the DUD library, indicating that the search algorithm of Autodock 4 isn't efficient in assessing ligands with multiple rotating bonds. [36] In addition, Vina can ran the benchmark 62 times faster than the Autodock and ran 7.25 times faster via the use of eight cores on the test machine. [36]

4b. Local Optimization of AutoDOCK Vina

By using L-BFGS, both the result obtained from the score function and the gradient of it is estimated. In terms of molecular docking, the gradient is defined as the first derivative of the scoring function with respect to the ligand information, which is location, orientation and the values of the torsions for the active notable bonds of the ligand and the residues nearby. In LBFGS, only the information from the number of iteration defined by the user and only this portion of the information is updated and stored in the approximated inverse Hessian matrix. When hitting local minima of the energy, the iteration will stop. Since the number of ligand in the database is usually large, L-BFGS can run faster than the regular BFGS, especially when the function used is non-linear. [36]- [40]

In Autodock Vina, the scoring function used is the combination of the empirical- and knowledge-based function. [36] It not only takes out the information from the preferential conformation of the ligand-receptor complexes but also the estimated binding affinity from experiment. [36] In general, the scoring function consists of the sum of the product between a series of interaction functions of the atom pairs and their distance, displayed in eq.1 [36]:

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij}) \quad (1)$$

Where c is the estimated score of the binding free energy, f_{titj} stands for a series of interaction function between each atom pair defined as i and j and r_{ij} is the interatomic distance between each atom pair respectively. [36] Similarly, the estimated score of the binding energy can also be treated as the sum of the inter- and intra- molecular contribution denoted as c_{inter} and c_{intra} respectively in eq.2:

$$c=c_{inter}+c_{intra} \quad (2)$$

According to the L-BGFS algorithm, s can be either the score of the global minima of the energy or the score of the local minima of the energy curve. [36] When the conformation of ligand receptor-complex is in the lowest energy form, s_1 is calculated as below:

$$s_1 = g(c_1 - c_{intra1}) = g(c_{inter1}) \quad (3)$$

In eq.3, g is the ascending smooth non-linear function, which was determined in eq.4 [36]:

$$g(c_{inter}) = \frac{c_{inter}}{1+wN_{rot}} \quad (4)$$

Where w is the coefficient of the number of active rotatable bonds between the heavy atoms in the ligand, denoted as $N_{rot.}$ [7] For the other conformations, only the c_{intra} of the best binding mode is used to estimate the binding free energy, which is denoted as s_i [36]:

$$s_i = g(c_i - c_{intra1})(5)$$

The scoring function listed in Eq.1 is also identical to the product between weighted sum of steric interaction $h_{t_i t_j}$ produced by each term adapted from Trot. et al [36] in Table1. below and the surface distance d_{ij} calculated from the difference between the interatomic distance and the Van der Waals radius of the two atoms [36]:

$$f_{t_i t_j(r_{ij})} \equiv h_{t_i t_j}(d_{ij}) \text{ (6.2)}$$

$$d_{ij} = r_{ij} - R_{t_i} - R_{t_j}(6.1)$$

In eq.6.1 R_{t_i} and R_{t_j} stands for the van de Waals radius of the atom i and j of the atom type t.

Table 1. shows the terms of the scoring function in Autodock Vina,adapted from Trot.et al[36].

Weight	Term
−0.0356	$gauss_1$
−0.00516	$gauss_2$
0.840	Repulsion
−0.0351	Hydrophobic
−0.587	Hydrogen bonding
0.0585	N_{rot}

The $gauss_1$, $gauss_2$ and repulsion term in Table1. are obtained by the following equations [36]:

$$gauss_1(L)=e^{-(L/0.5Å)} \text{ (7.1)}$$

$$gauss_2(L) = e^{-(\frac{L-3Å}{2Å})^2} \text{(7.2)}$$

$$\text{Repulsion}(L)=L^2, \text{ if } L<0 \text{ (7.3)}$$
$$=0, \text{ if } L \geq 0$$

$$\text{Hydrophobic}(L)=1, \text{ if } L<0 \text{ (7.4)}$$
$$=0, \text{ if } L \geq 0$$

$$\text{Hydrogen}(L)=1, \text{ if } L<0 \text{ (8)}$$
$$=0, \text{ if } L \geq 0$$

II. Biological Background of Death Associated Protein Kinase (DAPK2)

1. Structural Characterization of mDAPK2

Death-associated protein kinase 2 (mDAPK2) is a 370 a.a. CaM-dependent serine/threonine kinase of mouse death associated protein kinase (DAPK), a family of three proteins (mDAPK1, mDAPK2, and mDAPK3) whose kinase activity can be activated through calmodulin (CaM) binding under the stimulation of Ca^{2+} [41] [42]. Previous study has identified four functional regions for DAPK2: a catalytic kinase domain (3-274), an autoinhibitory region (292-301), a CaM-binding segment (287-354) (not shown), and a dimerization region (Figure A. and Sup. Figure A.)[41]. Similar to other DAPK family kinases, DAPK2's kinase domain contains a small N-terminal lobe and a large C terminus lobe, with the nucleotide binding cleft and the catalytic site in between. [41] At the N-terminus, a salt bridge is formed between the Glu64 in the active αC helix and Lys42 located at the active site of the DAPK2. [41] In the inactive conformation of DAPK2, this salt bridge is broken in one of two monomers, leading to the structural difference of the two active sites in the presence of bound nucleotides. (Figure B. and Sup. Figure B.) [41] At the C-terminus, the autoinhibitory region of DAPK2 consists of a linker region (residue 277-291) and the autoinhibitory helix αR1 (residue 292-301) (see Figure C. and Sup. Figure C.). [41] The autoinhibitory helix αR1 regulates the hydrophobic and electrostatic interactions between the peptide-binding site and the kinase core. [41] Recent study suggest that Lys297 and Lys298 located on the outside surface of the autoinhibitory helix is essential for CaM recognition while the

Phe296 and Tyr300 function as the two hydrophobic anchors of the autoinhibitory helix.

[41] (See Figure D. and Sup. Figure D.)

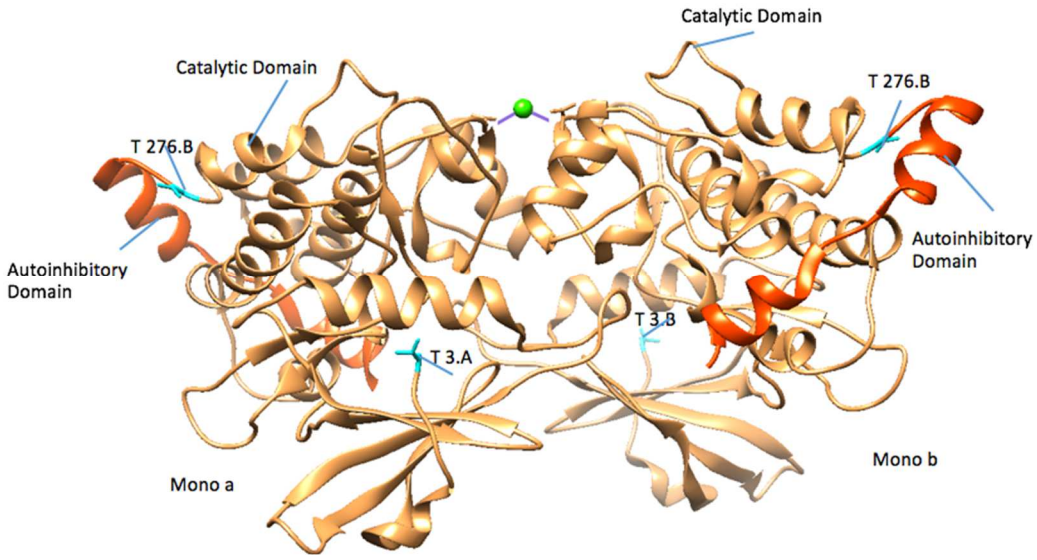


Figure A. shows the structure of the mDAPK2. The catalytic region is colored in sandy brown, a CAM binding domain (not shown, only the calcium ion is displayed in green), the autoinhibitory region is colored in orange red and the residues marking as the beginning and end of the catalytic domain (T.3 and T.276) are colored in cyan.

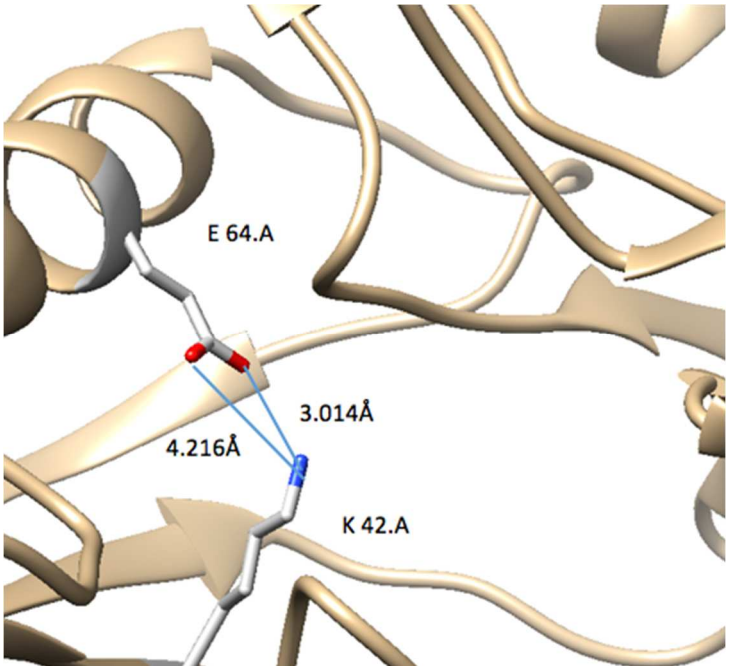


Figure B. shows the salt bridge between E64 and K42 (in light grey) at the active site of DAPK2**,** with the distance between side chain oxygen of E64 and nitrogen of K42 labeled.

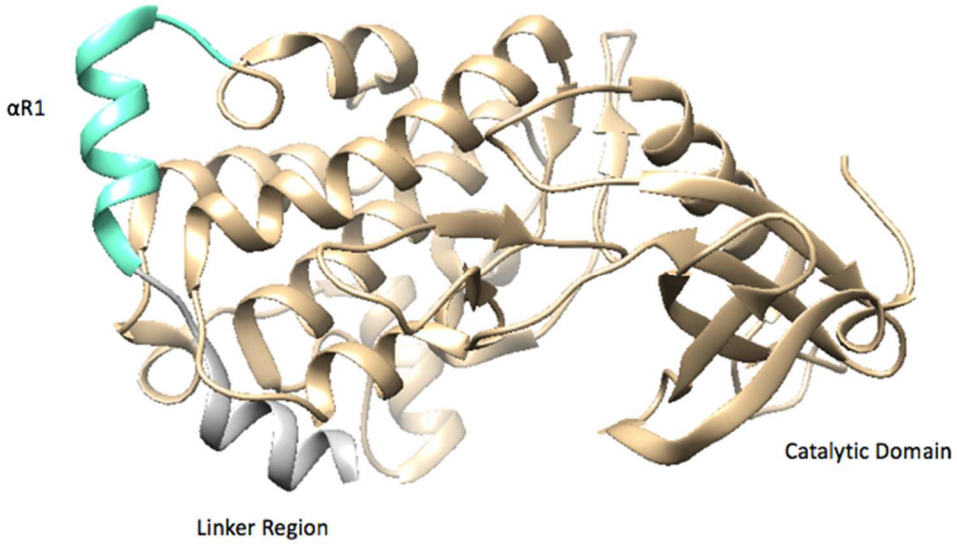


Figure C. shows the catalytic region (residue 3-274, in tan), autoinhibitory region of mDAPK2 consisting of a linker region (residue 277-291, in aquamarine,) and the autoinhibitory helix αR1 (residue 292-301, in gray).

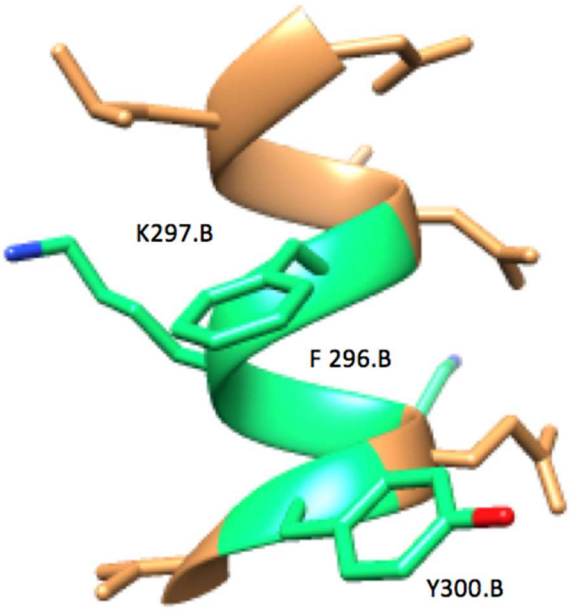


Figure D. shows the basic residue and hydrophobic residues located on the autoinhibitory helix (in spring green) of mDAPK2.

The sequence of DAPK family proteins shows overall high similarity among the three mouse kinases and their human homologs. The hDAPK1 shared 80.2%, 79.5%, 96.8%, 78.2% and 80.4% sequence identity with hDAPK2, hDAPK3, mDAPK1, mDAPK2 and mDAPK3 respectively in the protein kinase domain (Figure E.). Although the structural alignment between mDAPK2 and hDAPK2 indicates the highly structural similarity (Figure F.), the structural difference is surprisingly significant between hDAPK1 and mDAPK2 (Figure K.). The catalytic domain in both DAPK1 and 2 was found to be responsible for the dimerization process for both kinases, via the unique basic loop. [41] In mDAPK2, the basic loop b is inserted into a groove between helices αG and αR1, and hydrogen bonding interactions were formed through Residue47, 50 and 53. [41] (See Figure G. and Sup. Figure D.) Previous research also shows that the active site (see Figure H. and Sup. Figure E.) of the hDAPK1 is near the P-loop, which is located

between the basic loop (residue 45-56) and the activating loop. [41] [43] [44] P-loop is a glycine-rich region, which controls the catalytic activity of the DAPK1. When GTP binds to the P-loop, the catalytic activity of DAPK is reduced. [45] [46]

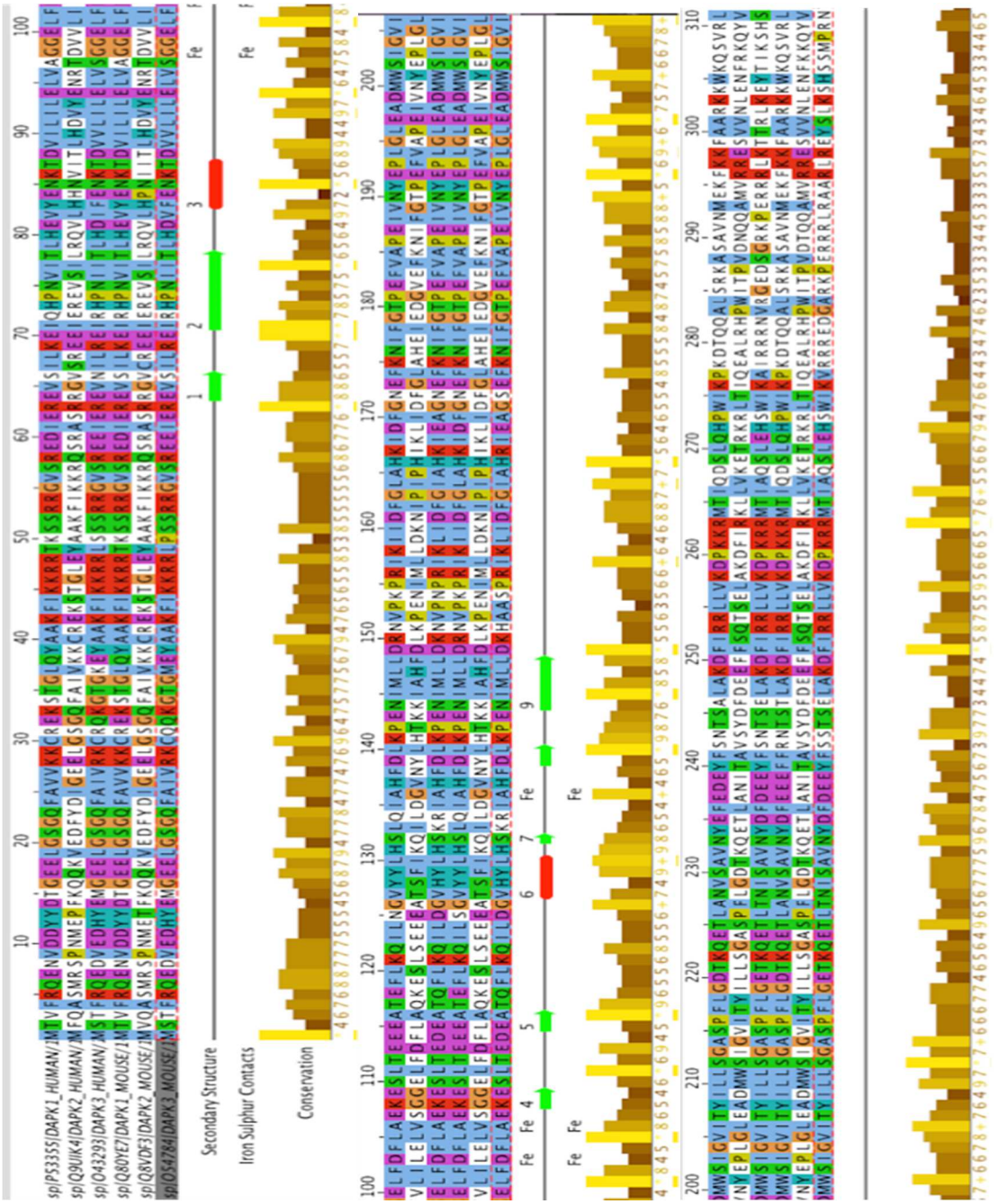
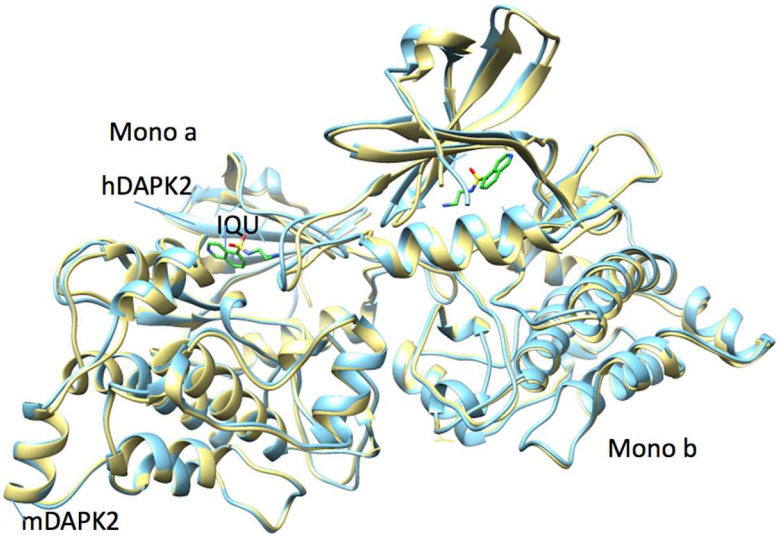
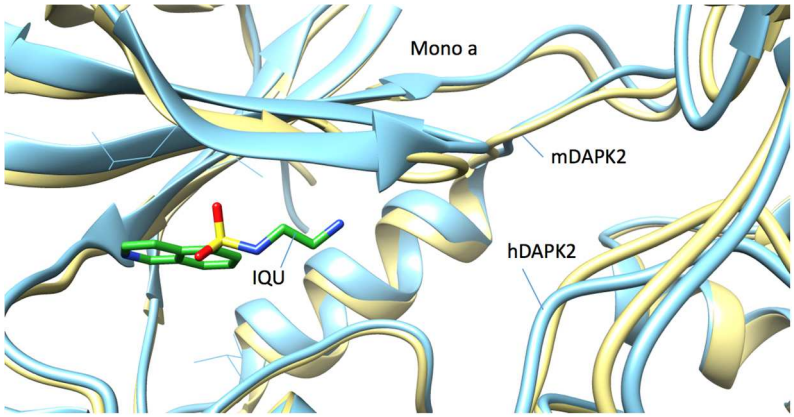


Figure E. Alignment of the catalytic domain of hDAPK1 with hDAPK2, hDAPK3, mDAPK1, mDAPK2 and mDAPK3. Asterisks indicate the conserved, amino acid residues within the kinase domain. The solid mono-color backgrounds indicate identical amino acid residues.

(a)



(b)



(c)

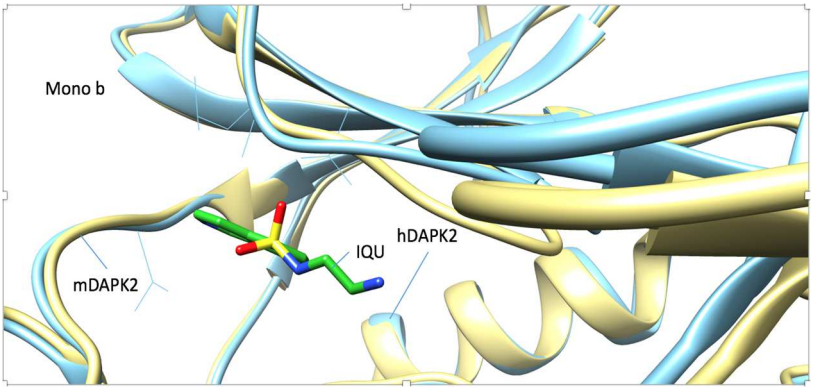


Figure F. shows the structural alignment of (a) the whole mDAPK2 (yellow)-hDAPK2 (blue), (b) the pocket view of the monomer a (ligand in green) and (c) b of DAPK2 in heterodimer form.

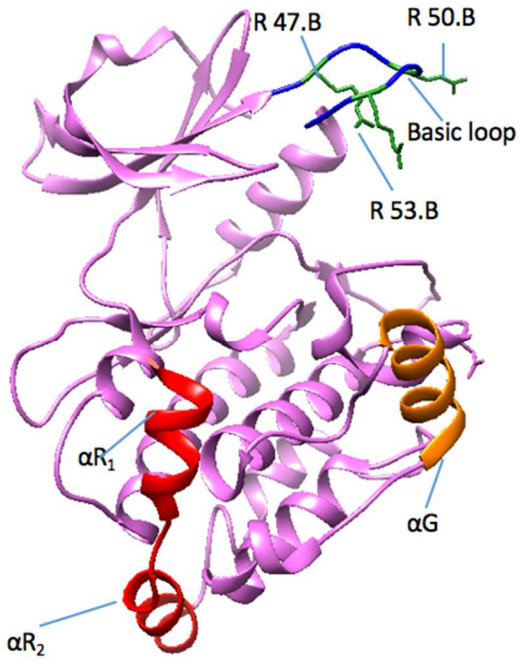


Figure G. shows the structure of basic loop (in blue) and the basic residue Arg47, Arg50, and Arg53 (in green), between helices αG (in orange) and αR1 (in red).

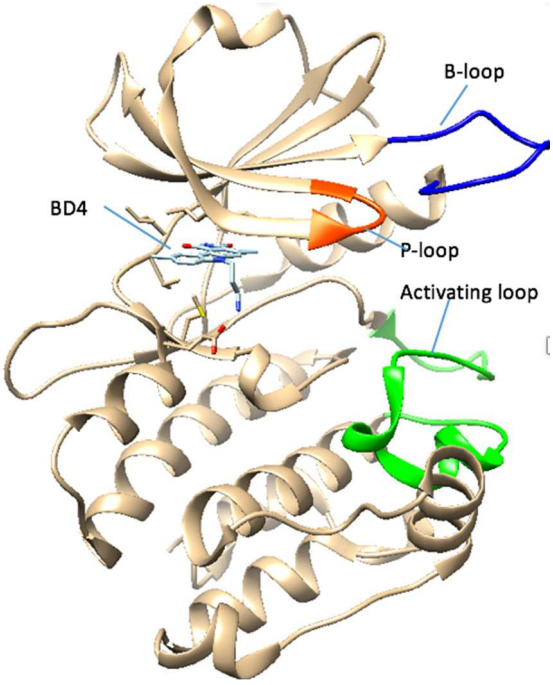


Figure H. displays the structure of the p-loop in hDAPK1 (in blue), located between the basic loop (in orange red), and the activation loop (in green).

2. Inhibitors Altering the Dimeric Autoinhibited Conformation of DAPK2

Compared with DAPK1 and 3, so far very few virtual screening researches has been conducted on DAPK2, although their structure is highly similar to each other. [41] [43] [44]**.** hDAPK1 has the largest number of specific ligand candidates, which is 13. hDAPK3 ranks the second, which is 4. hDAPK2 only has one specific ligand. Based on the previous studies, ligands binding specifically to hDAPK1, hDAPK2 and hDAPK3 were extracted and listed in Figures I. 1-2. Since DAPK1, 2 and 3 are structurally similar to each other; their corresponding ligand candidates also have comparable molecular structures and binding modes at the active sites. In Figure I. 3 M1 and M2 were candidates that displayed large % inhibition for DAPK1 and 3. [43] [44]

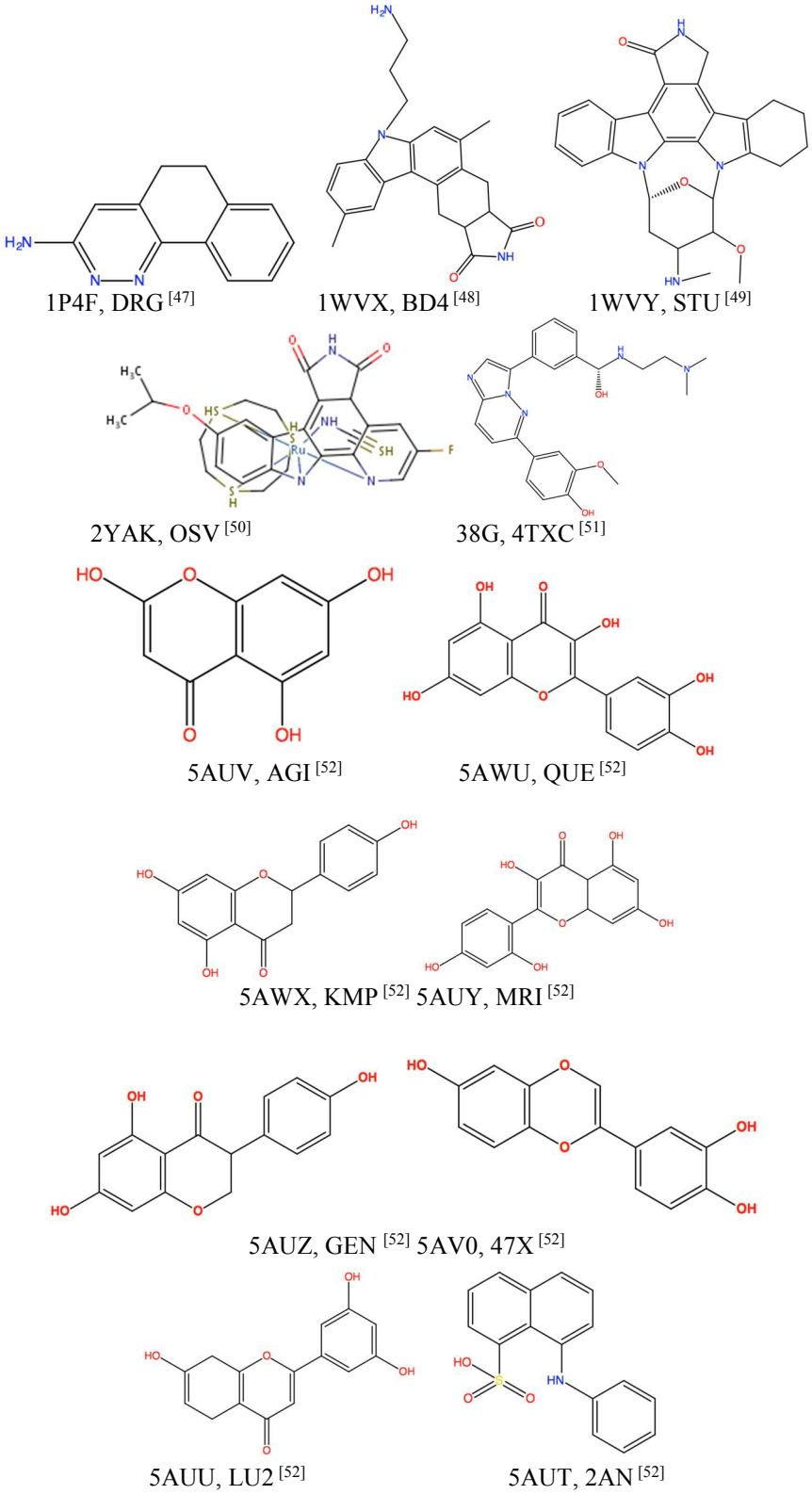
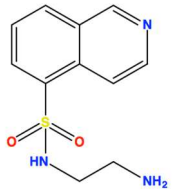
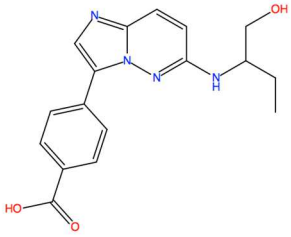


Figure I. 1. shows the molecular structures of ligands for (1) hDAPK: DRG (1P4F), BD4 (1WVX), 1WVY(STU), OSV (2YAK), 38G(4TXC), LU2 (5AUU) 2AN(5AT), AGI (5AUV), QUE (5AUW), KMP (5AUX), MRI (5AUV), GEN (5AUZ), 47X(5AV0), cont.

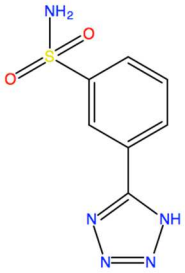


2CKE, IQU [44]

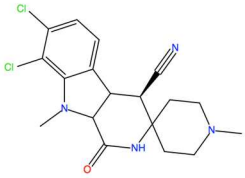
2J90, IZA [53]



3BHY, 4RB [54]

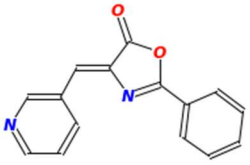


5A6N, U7E [55]

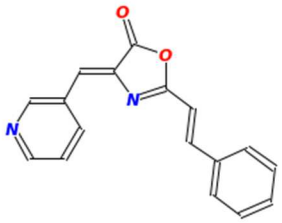


3BQR, 7CP [56]

Figure I. 2. shows the molecular structures of ligands for hDAPK2: IQU (2CKE) and hDAPK3: IZA (2J90), BR4 (3BHY), U7E (5A6N), 3BQR(7CP), cont.



M1. 84% inhibition at 10μM for DAPK1 and 100% inhibition 10μM for DAPK3



M2. 91% inhibition at 10μM for DAPK1 and 71% inhibition at 10μM for DAPK3

Figure I. 3. shows the molecular structures of M1 and M2, with the their %inhibition at 10μM of hDAPK1 and hDAPK3 labeled, cont.

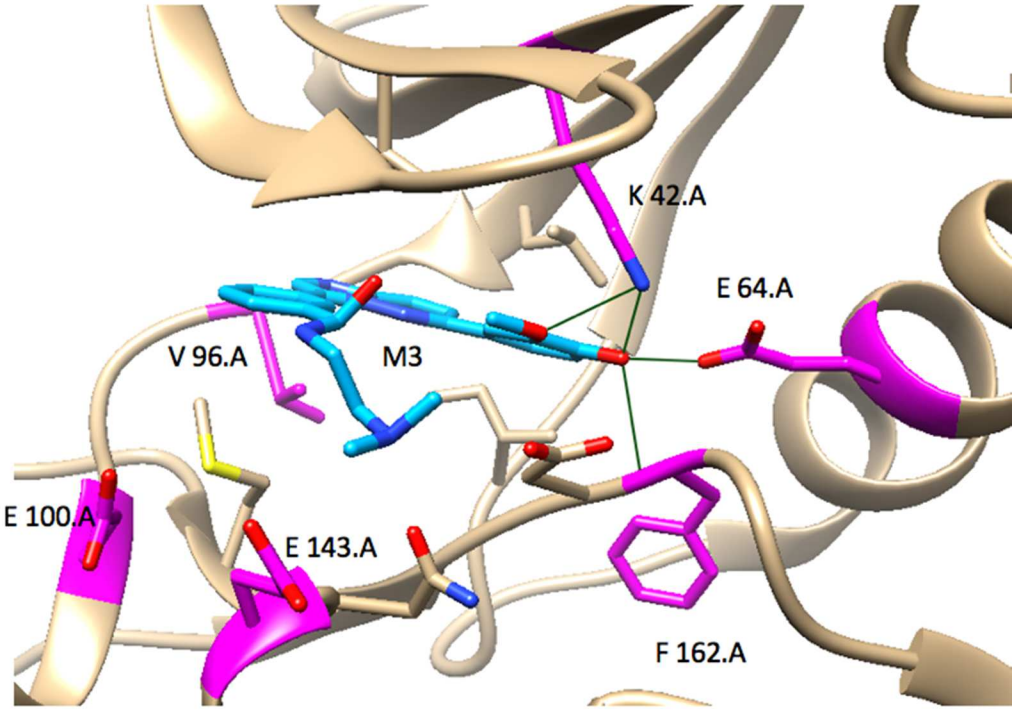


Figure J. shows the structure of the active site for M3 binding to DAPK1 at the active site, along with the pocket residues around the ligand. The hydrogen bond between the ligands and the residues were colored in forest green.

In Okamoto et. al, M1 is tested to be the most potent inhibitor for DAPK1 and 3. [43] Based on the predicted binding mode obtained from the CONSENSUS-DOCK, when M1 is docked into the active site of DAPK1, the N in the pyridinyl group of M1 forms a hydrogen bond with the NH backbone of VAL96. In addition, hydrophobic residues (Ala40, Met146 and Ile160) also involved in binding interaction between M1 and DAPK1. [43] M2 is a molecule whose structure is similar to M2, with a binding affinity stronger than M1. However, unlike M1 its binding mode still remains unknown. [423

In Wilbek et.al, M3 is identified to be another potent inhibitor for DAPK1 via molecule-screening method, with IC_{50}=0.247µm. [44] Looking at the binding pocket, the heterocyclic part of M3 forms a hydrogen bond between VAL96, with a fashion similar to the hydrogen bond formation between M1 and DAPK [41] [43] [44] (See Sup. Figure H.).

The hydroxyl-methoxy-phenyl portion of M3 also forms hydrophobic interaction with Lys 42 and Glu64. Currently, the crystal structure of DAPK1-M3 complex has an X-ray co-crystal structure at 1.9Å, with all of the residues clearly visible in the structure. [44]

Based on the previous research of the inhibitors for DAPK kinase [41] [43] [44] and the deviation of side chain for DAPK1 pocket residues comparing with the ones at DAPK2 (Figure K.), it's possible to find out potential inhibitors specifically for DAPK2. In Figure L., the pocket residues of DAPK1 listed in Okamoto.et al and Wilbek et.al are conserved in DAPK2. However, the side chain of these residues is not aligned in the same fashion. The side chain of residue 42,96,100,143,144 and 146 are deviated from each other, which might potentially lead to the difference in ligand selection. In this project, the binding pocket of DAPK2 was identified based on the alignment results in Figure K. and Figure L.

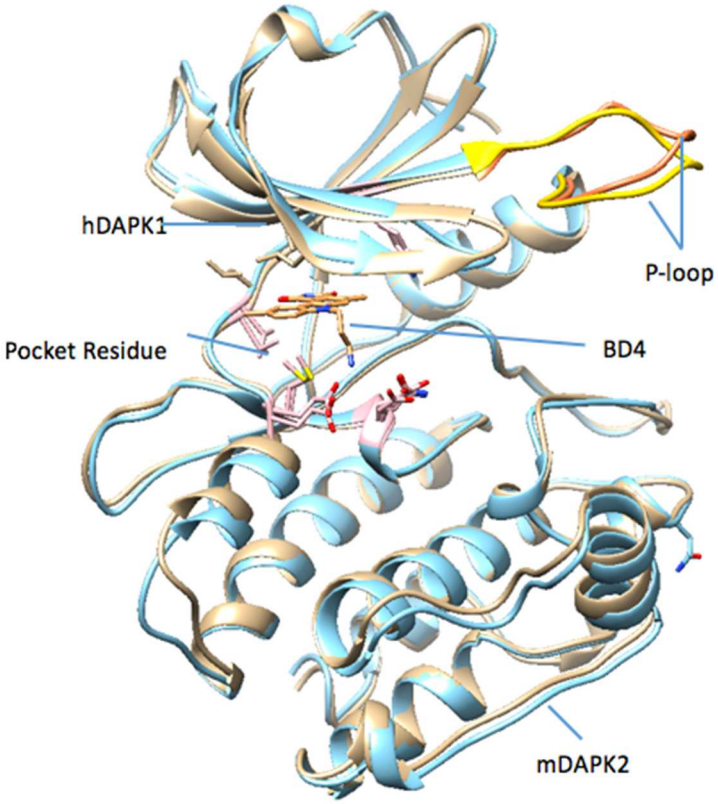


Figure K. shows the whole view of the aligned hDAPK1 and mDAPK2. The P-loop of DAPK1 and 2 is colored in coral and gold respectively, and the residues colored in pink are the pocket residues at the binding pocket.

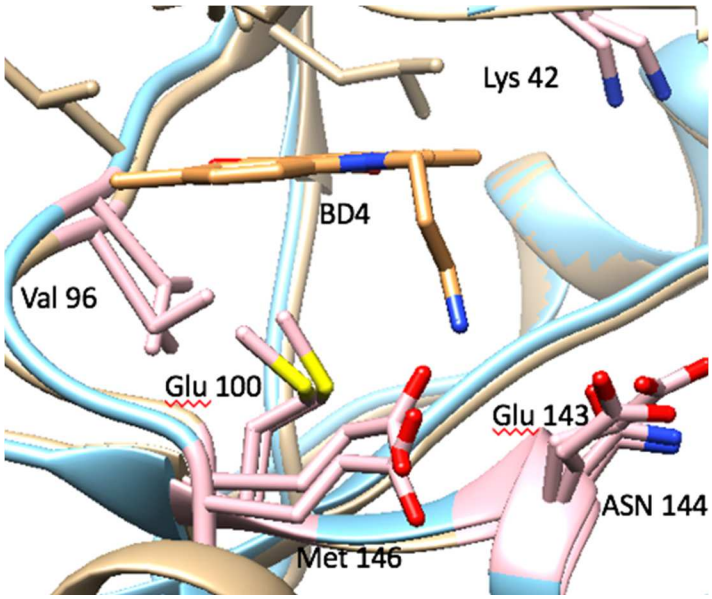


Figure L. shows the structural view of the binding pocket for hDAPK1, with mDAPK2 aligned together. Residue 42, 96, 100, 143, 144 and 146 are colored in pink and with the side chain displayed, for both DAPK1 and 2.

3. Medical Motivation of Screening Inhibitors for DAPK2

Although the biological function of DAPK2 is closely related to leukemia and the lymphatic tumor, it's not an ideal target drug for cancer due to the insufficient knowledge background of DAPK2. [57] As a result, the drug discovery of specific activators of genes downregulated by promoter hypermethylation is restricted. [57]

More recently, it has been reported that DAPK1 is the potential drug target for diseases characterized by neurodegeneration, such as stroke or traumatic brain injury [57][58] In PELLED D, RAVEH T, RIEBELING C et al, DAPK protein expression was unregulated in rat hippocampal neurons after1 h of exposure to ceramide and cell death occurs after 16 h of exposure, indicating that DAPK might involve in the apoptosis of neuronal cell via the interaction between the catalytic activity and the death domain interaction [59]. Since DAPK1 and 2 is highly similar to each other structurally, it's possible that DAPK2 might also be a potential target similar to DAPK1. As a result, discovering new inhibitors for DAPK2 regarding the prevention of neurodegenerative diseases should be of importance of such endeavors.

III. Biological Background of Engulfment and Cell Motility 1 Protein (ELMO1)-Dedicator of Cytokinesis 2 (DOCK2) Complex

1. Structural Characterization of ELMO1

ELMO1 stands for engulfment and cell motility 1 protein, which belongs to the Engulfment and Cell Motility family (ELMO). It's a human protein regulated by ELMO1 gene with 720 residues in total. [60] [61] It has a ras homology gene family at the N terminus, a Rho-G binding region, the ELMO domain, the PH domain and a sequence with three Pxxp motifs at the C-terminus. (See Figure N. 1.) [62] The ELMO domain binds to the

SH3 domain of DOCK2 at its C-terminus, then the SH3•DHR2 interaction and in this way the DOCK2 protein can activate the Rac GTPase. [62] [63] When ELMO1 binds to DOCK2, an intermolecular five-helix bundle can also form by the alpha helical region, the SH3 domain, and the extensions of the PH domain of ELMO1 [61]. Due to the interaction of the ELMO inhibitory domain (EID) with the ELMO auto-regulatory domain (EAD), ELMO1 is reported auto-inhibited. [61] However, DOCK2 and ELMO1 together can relieve their auto-inhibitions, whose mechanism still remains unknown. [61]

2. Structural Characterization of DOCK2

Dedicator of cytokinesis 2 (DOCK2) is a guanine nucleotide exchange factor carrying 168 residues with a mass of 213 kDa. [61] It's a Rac activator that manages the motility and polarity in neutrophil chemotaxis. [63] DOCK2 is a member of CDM family (Caenorhabditis elegans CED 5 mammalian DOCK180, and Drosophila Melanogaster Myoblast city), and shares sequences highly similar to DOCK180. [61] Like 10 other DOCK proteins in the CDM family, DOCK2 also has a DOCK-homology region (DHR-1) and DHR-2 but doesn't have a DH-PH domain. [61] DOCK2 facilitates the Rac GEF reaction via DHR-2 domain. [61] It also binds to the phosphatidylinositol 3,4,5-triphosphate (PIP_3) via DHR-1 domain, which is required for the interaction between DOCK2 and the plasma membrane. [61] At the N terminus, DOCK2 contains a SH3 where the ELMO1 interacts. [63] Moreover, the SH3 domain of DOCK2 can bind to the C-terminus of ELMO1 not just in the PxxP region. [61] When ELMO1 binds to the DOCK2, the autoinhibition of DOCK2 can be relived. [61]- [63]

3. Interactions that Affect the Binding Between ELMO1 and DOCK2

According to K. Namekata et.al and Hanawa-Suetsugu et.al, ELMO1 can bind to DOCK2 via the the interaction between the Pxxp domain at ELMO1 and SH3 domain at DOCK2, which lead to the Rac access to the catalytic DHR-2 domain. [61] [64] In addition, Hanawa-Suetsugu et.al also found that a five helix bundle is formed by Dα1-3 of DOCK2 and Eα1 and 3 of ELMO1, which involves a large number of aliphatic side chains of ILE and Leu residues. [61]

At the C-terminus of ELMO1, Pro 707,710,711,712,714 and 717 together form three PxxP motifs. [61] The ELMO1 peptide binds to the DOCK2 at the SH3 domain and form a polyproline II helix. [61] Tyr 44 of DOCK2 binds hydrophobically with Pro714 and Pro717 of ELMO1, meaning that Pro174-x-x-Pro717 is the PxxP motifs that DOCK2 binds at (Sup. Figure K.). [61][64]] The Tyr 44 of DOCK2 also forms hydrogen bond with Lys 715 at the PxxP motif of ELMO1. [61] When interacting with each other, DOCK2 and ELMO1 forms a large, rigid complex involving the entire regions of the two protein fragments. [61]

The intermolecular five-helix bundle is constituted by Dα1-3 of DOCK2 and Eα1 and 3 of ELMO1 (Sup. Figure L.). [61] From the previous structural analysis, Leu 95, Trp96, Trp 102, Val 107, Phe 114, Met 121 and Met 125 at DOCK2 interact with the Ile 544, Leu 547, Ile 548, Leu 681, Leu 689, Met 692, Leu 696, and Leu 699 at ELMO1 (Sup. Figure M.). [61] Besides hydrophobic interactions, several hydrophilic interactions also play an important role in the formation of five-helix bundle. The side chain of

Lys103, Tyr 106 and Arg 128 of ELMO1 form hydrogen bond with the side chain of Glu693, Arg 685 and Leu 699 of DOCK2 respectively (Sup. Figure N.). [61]

Based on the crystal structure of the ELMO1-DOCK2 complex, it is possible to find out inhibitors that can interrupt the formation of the complex, via binding to the pocket located at either SH3 domain of DOCK2 or the six bundle helices formed between ELMO1-DOCK2.

4. Finding out Inhibitors that Could Interrupt the Binding between ELMO1-DOCK2 Complex

ELMO1 plays an important role in cell migration. [64] [65] ELMO1 can act downstream of the phosphatidylserine receptor BAI and activate RAC1 via the formation of complex with DOCK1, which is responsible for the promotion of cell motility and engulfment. [60] [66] [67] The activation of RAC1 via formation of ELMO1-DOCK complex was found to involve in: the internalization of apoptotic cells leading to serious testicular problems [68], increase of lymphocyte migration of primary T cells [69], and actin cytoskeleton regulation during breast cancer. [70] In addition, in John W. et al. [65] ELMO1 is found sufficiently in the inflamed synovium and the inhibition of ELMO1 transcription with siRNA in rheumatoid arthritis fibroblast-like synoviocytes (RA FLS) can result in the reduction of cell migration and invasion. [65] Since Rac1 GTPase activity was regulated by ELMO1; John W. et al. also indicated that ELMO1 could participate in the cell motility and invasion when Rac1 is activated. [65] In Lefevre et.al and Lee et. al, FLS is responsible for the cartilage damage in RA and can potentially migrate to other joints. [71] Therefore the blockage of migration can slow down the progress of disease. [65] [71] [72]

When ELMO1 complexes with DOCK2, similarly Rac GTPase can also be activated and thus the cell migration can be increased. [61] From the structural characterization of ELMO1-DOCK2, there might be potential pockets located at the interface between ELMO1 and DOCK2, where the inhibitors can bind. The RA FLS migration by Rac1 can be prohibited potentially, via the disruption of ELMO1-DOCK2 complex by the inhibitors. [61] [65] [71] [72] Although RA drug development is still a challenge; RA can be treated potentially via prohibiting the formation of ELMO1-DOCk2 complex.

Chapter 2 Method and Results

I. Identification of the Binding Pocket

1. Pocket Identification on DAPK2 via Structural and Sequence Alignment

Identification of binding pocket for mDAPK2 is essential prior to docking. However, mDAPK2's pocket residues cannot be determined directly from the structural information due to the lack of any inhibitor binding complex structure. Therefore, we used the inhibitor binding structures from human DAPKs to identified the pocket residues of mDAPK2 since they share very high similarity on sequence, for ex. hDAPK1 and mDAPK2 both have a catalytic domain that share 80% identity at the amino acid level. First, the multiple sequence alignment on the catalytic domain was performed between hDAPKs (PDBID: 1WVX, 2CKE, 3BHY) and mDAPK2 (PDBID: 2YA9) using ClustalW. The quality of the multiple sequence alignment is high due to the overall high similarity between the sequences of DAPK proteins. No gap is observed on the kinase catalytic domain. Next, the pocket residues are identified from each of the structure using the criterion of minimum distance less than 8Å to the ligand. The pocket residues from each complex structure are mapped to the multiple sequence alignment. The union of these pocket residues are selected as the pocket residues for mDAPK2, as displayed in Figure M. The mDAPK2 pocket was finally constituted by the following residues: L19, G20, S21, K42, L93, E94, L95, V96, S97, G98, G99, E100, P142, E143, N144, I145, and D161.

Figure M. The result of sequence alignment between hDAPK1 (1WVX), hDAPK2 (2CKE), hDAPK3 (3BHY) and mDAPK2 (2YA9), via Jalview [73]. The residues within 8Å of the ligands in each DAPK receptor were grouped and colored accordingly.

2. Prediction of Potential Binding Pocket for ELMO1-DOCK2 Complex via LIGSITE^{csc}

According to the present study [61] [65], it is possible to find out small-molecule inhibitors that can disrupt the formation of ELMO1-DOCK2 complex at either the SH3 domain or between the five-helix bundles. To test our hypothesis regarding the binding pockets of the ELMO1-DOCK2 complex, LIGSITE^{csc} is used, where superscript csc stands for Connolly surface and conservation. It's a software designed exclusively for ligand binding site, based on the surface-solvent-surface events and the conservation of the involved surface residues. [74] Instead of using the x, y, and z coordinates of the protein atoms to locate the binding site, the protein's solvent accessible surface is used.

Nowadays, lots of software such as Pocket, LIGSITE, SURFNET, CAST, and PASS are designed to predict and analyze the potential binding site for ligands, via characterizing the geometric features of the receptor merely. [74] In Huang and Michael et.al, LIGSITE^{csc} was compared with the previous developed Pocket, LIGSITE,

SURFNET, CAST, and PASS over 48 unbound/bound and 210 bound-only receptor-protein complex by the same method of assessment. [74] LIGSITE^{csc} can identify the ligand-binding site correctly in 71% and 75% cases respectively, which are better than the other software mentioned here. [74]

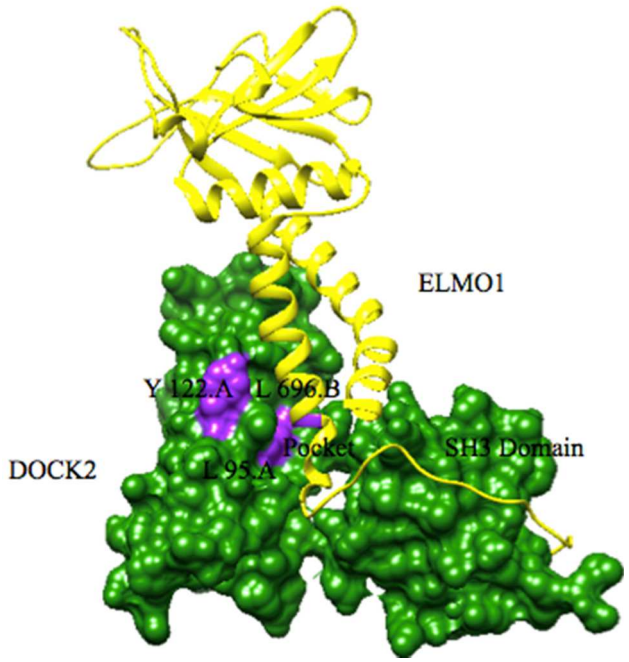
Here is how the LIGSITE^{csc} works [74]: First of all, the protein, which was ELMO1-DOCK2 complex in this article, was traversed into a 3D grid with a step size of 1Å. In this step, the x, y, and z components as well as the four cubic diagonal vectors were analyzed. After scanning the protein complex in 7 directions, grid points were marked as protein, surface or solvent using the following criteria (Sup. Figure O.) [75]: If a grid point is located within 1.6Å of the atom, then it is marked as protein. To confirm protein surface, Connolly algorithm is applied. A virtual probe with a defined size, which is 1.4Å in LIGSITE^{csc}, is used to orbit the protein. As a result, a solvent accessible surface consisting of the Van der Waals surface of protein and probe surface is formed followed by the storage of the coordinates of the surface vertices (Sup. Figure P.) [75]. If the distance between the vertices of solvent accessible surface and a grid is less than 1Å, and if this grid point meets the condition to be label as protein, then this grid point is labeled as surface. All other points that don't fall into any of the criteria described above are labeled as solvent. At the reentrant surface, lines were drawn between two surface points, which passed several solvent grids. The combination of surface grids and the solvent grids here is called surface-solvent-surface event. Excluding the surface grid, if the number of solvent grid in between is larger than six, then these solvent grids are marked as pocket. Then these pocket grids were clustered according to the proximity and the clusters was ranked according to the number of pocket grids. Finally, these pockets

were ranked again based on the degree of conservation of the surface residues that located within 8Å of the cluster center. [74] [76]

In this project, both the structural-based and ligand-based virtual screening are performed at the binding pockets predicted by LIGSITE^{csc}, to search for ligands disrupting the formation of ELMO1-DOCK2 complex.

To find out if there're any potential binding pocketed between the 5-bundle helix structure between ELMO1 and DOCK2, LIGSITE^{csc} was used. According to Hanawa et.al [61], the prediction of binding sites was performed on chain A and B only. Figure N. shows the 2 most potential binding sites for ELMO1-DOCK2 complex, by setting the number of potential binding sites to 3.

(1)



(2)

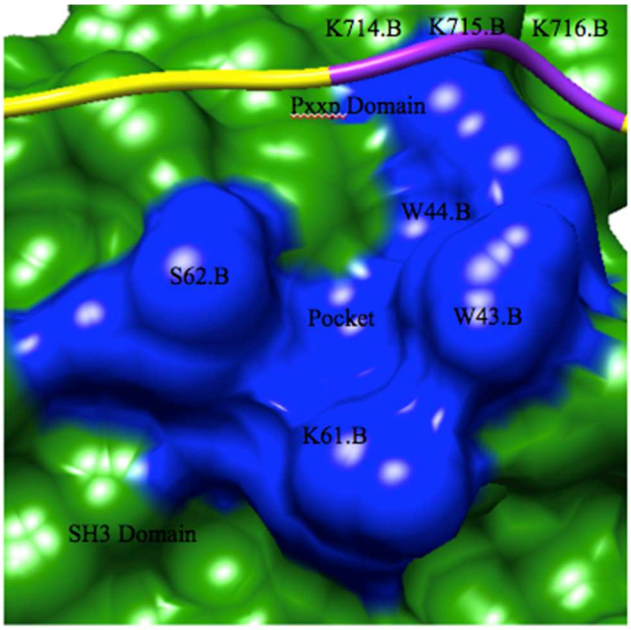


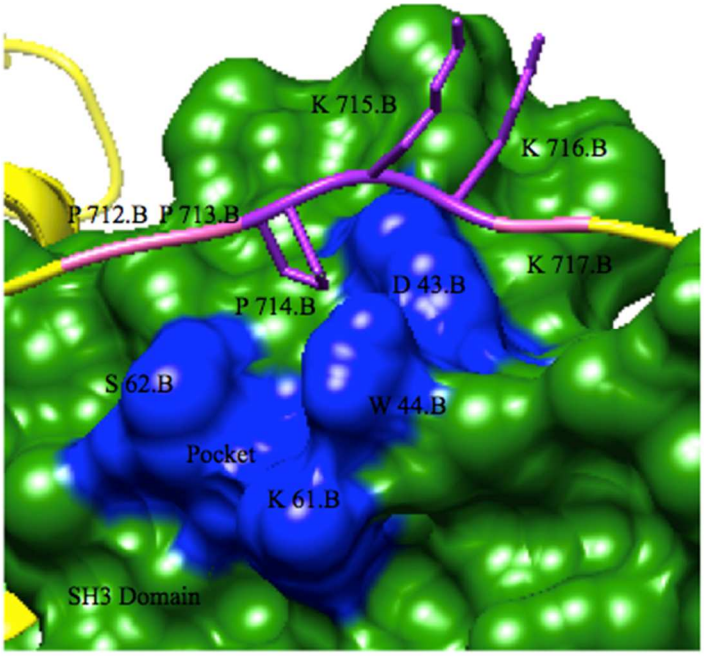
Figure N. 1-2. Two of the most likely binding site for ELMO1-DOCK2 complex (PDBID: 3A98).
Figure N. 1. The binding site at the five-bundle helices between ELMO1 (colored in yellow) and DOCK2 (colored in green) consists of Y122.A, L96.A and L696.B, cont.

Figure N. 2. The other binding site for ELMO1-DOCK2 complex consists of two domains: SH3 and Pxxp domain. At the SH3 domain of DOCK2, these residues are the ones that form the binding site (colored in blue): Asp 43, Trp 44, Lys61 and Ser62 (Inglis et al. and N.Atatrech et al.). At the Pxxp domain, these residues are the ones that form the binding site Pro 714, Lys 715 and Glu 716 (Hanawa-Suetsugu et al.).

To figure out the center coordinates and the size of the search volume for the binding pocket at the SH3 domain, residues that constitute the binding site were selected, as it is shown in Figure O. 1 and Figure O. 2. In addition to these residues, two residues at each end of the Pxxp domain (Pro 714, Lys 715 and Glu 716) were also selected respectively. By using the x, y, and z coordinates of these selected 10 residues in Figure O. 1, the size of the length, width and height of the search space was 19.954, 22.208 and 21.670 respectively. Finally, the information regarding center coordinates of the predicted binding pocket and the size of the search volume were saved as a configuration file, which would be used in virtual screening.

To confirm the center coordinates and the size of the search volume for the five-bundle helix site, residues within 8Å of L696.B were selected, as displayed in Figure N. 1. The reason why L696.B was selected as the center residue is that the ligand candidates should be able to insert into the 6-helix bundle and thus can interrupt the interaction between ELMO1 and DOCK2. Using the x, y and z coordinates of the 8Å residues in Figure O. 2 were extracted and used to calculate the center of the potential binding pockets and the size of the search space. The size of the length, width and height of the search space was 17.237,21.765 and 19.679 respectively. Finally, the information regarding center coordinates of the predicted binding pocket and the size of the search volume were saved as a configuration file, which would be used in virtual screening.

(1)



(2)

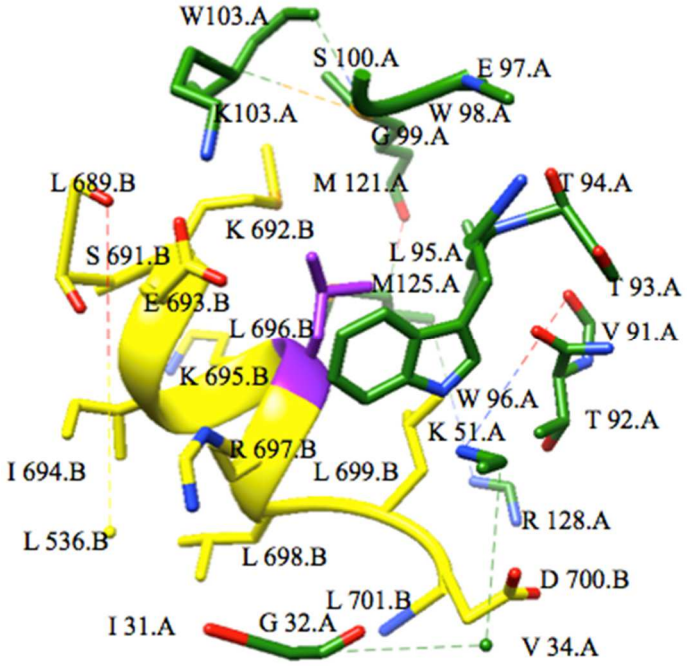


Figure O. 1-2. Residues that determines the center coordinates and search volume of ELMO1-DOCK2 complex.

Figure O. 1. In addition to residues labeled in Figure N-2, two residues (pink) at each end of the Pxxp domain (Pro 714, Lys 715 and Glu 716) were selected respectively, cont.

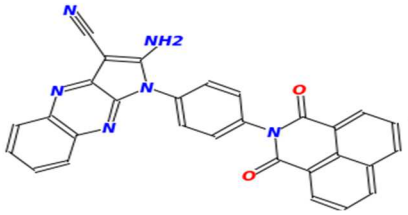
Figure O. 2. Residues within 8Å of L696.B (I31, G32, V34, K51, V91, T92, T93, T94, L95, W96, E97, W98, G99, S100, W102, K103, M121, M125, R128, L536, L689, L690, S691, M692, E693, I694, K695, L696, R697, L698, L699, D700, I701) were colored in purple.

II. Docking Inhibitors on the Previously Identified Pocket on DAPK2

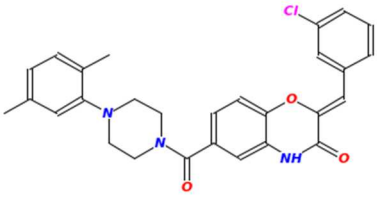
1. Docking on mDAPK2

In this article, all of the ligands in the database are in 3D structures, which can be obtained from the ZINC database. [77] The "drug-like" ligand subset was downloaded from ZINC database [77] containing 17,900,742 ligands. All the ligands were processed into pdbqt format using script prepare_ligand4.py provided in AutoDock Vina package. mDAPK2 structure was taken from PDB ID 2YA9 and processed into pdbqt format by script prepare_receptor4.py in Autodock Vina. [36] [78] Then the x,y and z value of center coordinates and the magnitudes of length, width and height of a docking cube are determined from the coordinates of the pocket residues of DAPK2. In rigid docking, the level of exhaustiveness, which tells how hard the Vina should perform the conformational search of ligands, was set to 1. [78] Only the very best docking conformation is kept for each ligand. [78] Part of the rigid docking results is shown in Sup. Figure G. and Sup. Table 1. [78]

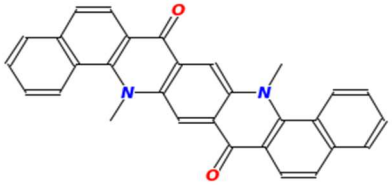
The top 1 million candidate ligands from rigid docking were selected for the following flexible docking. The same binding pocket setup is used for flexible docking as the rigid docking. The exhaustiveness, the binding mode and energy range was set to 9,1 and 3 respectively. Figure P. 1-10 shows the top 10 candidate ligands for DAPK2 from the flexible docking result. Based on their binding affinity difference produced by DAPK1 and 2, ZINC 32721460 was selected to perform MD simulation described in the next section. For the simplified version of the docking process, please refer to the flow chart in Sup. Figure J. 1.



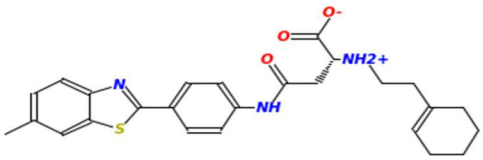
1.ZINC02655180
(-12.1Kcal/mol, 480.487g/mol)



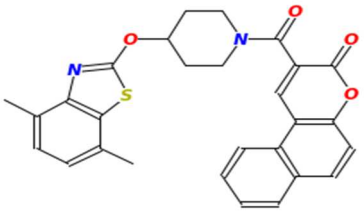
2. ZINC32983948
(-11.9Kcal/mol, 480.987g/mol)



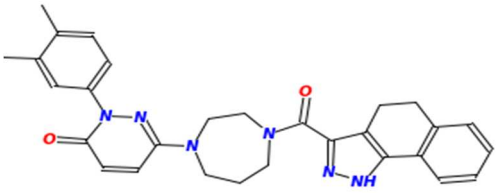
3. ZINC22007970
(-11.9Kcal/mol, 440.502g/mol)



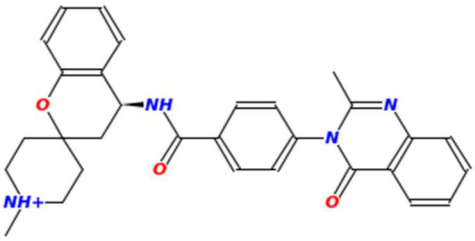
4. ZINC72438491
(-11.9Kcal/mol, 489.531g/mol)



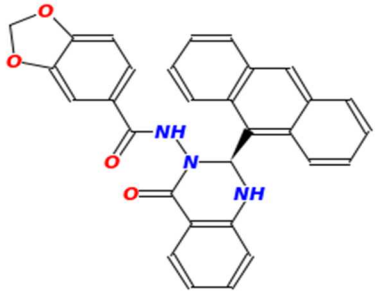
5.ZINC49442684
(-11.8Kcal/mol, 484.577g/mol)



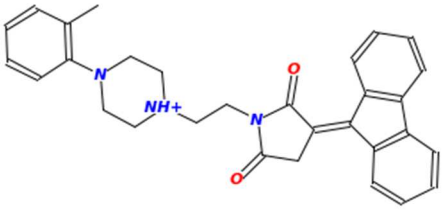
6. ZINC15730186
(-11.8Kcal/mol, 494.599g/mol)



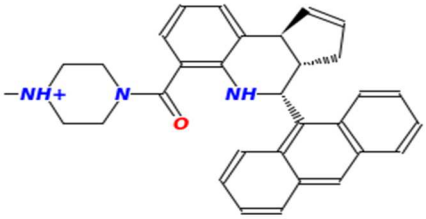
7. ZINC32721460
(-11.7Kcal/mol, 495.603g/mol)



8.ZINC00671505
(-11.7Kcal/mol, 487.515g/mol)

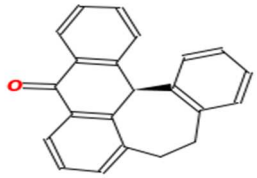


9.ZINC08399012
(-11.7Kcal/mol, 464.589g/mol)

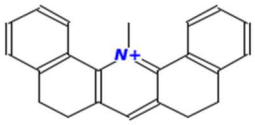


10. ZINC14549309
(-11.7Kcal/mol, 474.628 g/mol)

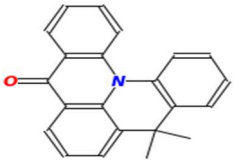
Figure P. 1-10. Top 10 ligand-candidates for Rigid Docking at DAPK2 (exhaustiveness=9). All of the ligands were ranked based on their binding affinity, numbered from 1 to 10.



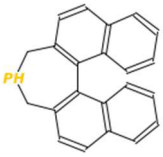
1. ZINC04760022
(27420,296.369g/mol, -10.0Kcal/mol)



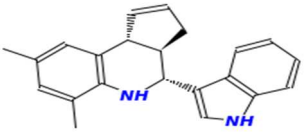
2. ZINC05377817
(37565, 298.409g/mol, -9.9Kcal/mol)



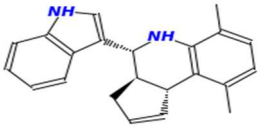
3. ZINC05482357
(33619, 311.384g/mol, -9.9Kcal/mol)



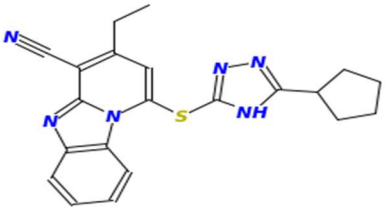
4. ZINC71772079
(47566,312.351g/mol, -9.8Kcal/mol)



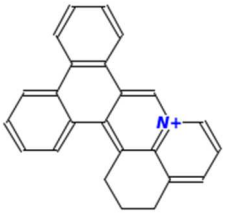
5. ZINC14544618
(32893,314.432 g/mol, -9.9Kcal/mol)



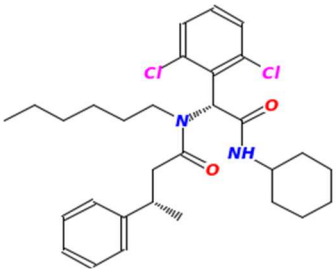
6. ZINC14544694
(33322, 314.432g/mol, -9.9Kcal/mol)



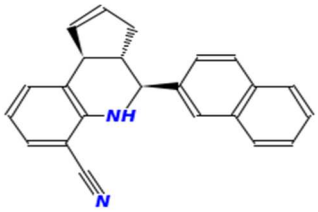
7. ZINC14321140
(14513,320.415g/mol, -10.2Kcal/mol)



8. ZINC01019606
(35596,314.473gmol, -9.9Kcal/mol)



9. ZINC36515879
(36349,321.335g/mol, -10.0Kcal/mol)



10. ZINC14475072
(27080,322.411g/mol, -10.0Kcal/mol)

Figure Q. 1-10. Top 10 candidates with the lowest molecular weight among candidates in DAPK2, ranked according to the molecular weight. The molecules were numbered according to their molecular weight.

2. Ligand Similarity of Flexible Docking Results

To estimate how similar the ligand candidates are for the DAPK2 receptor, the Tanimoto coefficient for 1000X1000 pairs of ligands was initially calculated using the path-based fingerprint FP2 in openBabel [79]. This fingerprint identifies the linear and ring structures of the molecules up to the length of 7, and these substructures were mapped via a hash function. [79] The results of Tanimoto coefficients for the 1000X1000 pairs of ligands were then saved in a single file. Tanimoto coefficient (TC) is the most widely used coefficient in ligand based virtual screening, calculated by eq.9:

$$T_c = \frac{c}{(a+b-c)} \quad (9)$$

Where c stands for the total number of ligands in both molecular fingerprint A and B, b stands for the number of ligands found exclusively at B and a stands for the ligands found exclusively in fingerprint A. [4] [80] The range of Tanimoto coefficient is between 0 and 1. [4] [80]

The top 1000 ligands from flexible docking result are used for similarity analysis. A 1000X1000 matrix was made by calculating the TC of each ligand pairs. The Tanimoto coefficient matrix was converted into distance matrix by subtraction of each element from 1. The distance matrix was clustered using the complete linkage hierarchical clustering algorithm. In each clustering step two nearest clusters are gathered and form a new cluster until the whole data set is merged into one single cluster. Complete linkage is the method to calculate the distance between two clusters that the distance between two clusters is defined by the maximum distance between the elements from the two clusters. The hclust function in R is used to perform the hierarchical clustering calculation The hierarchical clustering result was then divided into 731 clusters

using the TC cutoff 0.9 so that the ligands with TC more than 0.9 are clustered into one cluster. We concluded that the redundancy of the flexible docking result is not high since most of the clusters only contain one or two ligands. For the simplified version of the ligand similarity analysis method, please refer to the flow chart in Sup. Figure J. 2.

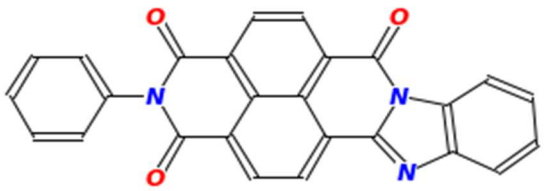
3. Binding Specificity between hDAPK1 and mDAPK2

The docking result of mDAPK2 shows that the majority of the inhibitors are rigid aromatic molecules, which are similar to previous identified hDAPK1 inhibitors. However, we also find that the potential inhibitors of mDAPK2 are much larger than that of hDAPK1. Such observation suggests that mDAPK2 and hDAPK1 could have different binding specificity. To investigate the binding specificity of the two kinases, we docked the top 300 thousand ligands from the flexible docking result of mDAPK2 to hDAPK1. The structure of hDAPK1 was taken from PDB ID 1WVX. The docking parameter of hDAPK1 used the flexible docking protocol of mDAPK2, where exhaustiveness, number of model, and energy range are set to 9, 1 and 3 respectively.

For each of the 300 thousand ligands, the difference of estimated binding energy is calculated between mDAPK2 and hDAPK1. We ranked the ligands using the difference and picked the top one and the bottom one representing the mDAPK2 preferred ligand (DA7, Fig P1) and the hDAPK1 preferred ligand (BAL, Figure P. 2. and Sup. Figure H. 3.) respectively, to look into their binding energy profile using the MM/GBSA method.

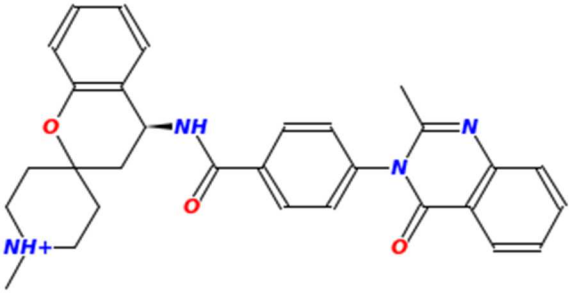
Figure R. 1-2. shows the Ligands selected to perform MD simulation on DAPK1 and 2 respectively. These ligands were selected based on the difference between their binding affinity and their ranking in DAPK1 and 2 respectively.

(1)



Name: PhenylBLAHtrione
ID: ZINC00988780
Rank in DAPK2: 31468
Rank in DAPK1: 5
Binding Affinity for DAPK1: - 12.3Kcal/mol
Binding Affinity for DAPK2: -9.9Kcal/mol
Difference: -2.4Kcal/mol

(2)



Name: 4-(2-methyl-4-oxo-quinazolin-3-yl)-N-[(4S)-1'-methylspiro[chromane-2,4'-piperidine]-4-yl]benzamide
ID: ZINC32721460
Rank in DAPK2: 7
Rank in DAPK1: 244853
Binding Affinity for DAPK1: -8.9Kcal/mol
Binding Affinity for DAPK2: -11.7Kcal/mol
Difference: 2.8Kcal/mol

Figure R. 1-2. shows the ligand candidates chosen to perform MD simulation on DAPK1 and 2. Ligand in E1 is DAPK2-specific ligand, while E2 is DAPK2-specific ligand. P1: Structure of ZINC00988780 (BAL). P2: Structure of ZINC32721460 (DA7).

In this section, molecular dynamics (MD) simulation was performed to track the physical movement between the ligand and the protein, after the potential candidates were selected from the virtual screening process. The whole MD simulation process was carried out by a MD simulation software named Assisted Model Building with Energy Refinement 15 (AMBER 15). [81] [82] The potential function used to calculate the total energy of the covalently bonded atom pairs for the whole complex is described below [82]:

$$E_{complex} = \sum_{angle} K_\theta(\theta - \theta_{eq}) + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\beta - \gamma) + \sum_{bonds} K_a(a - a_{eq}) + \sum_{i<j}[\frac{A_{ij}}{R_{ij}^{12}} + \frac{B_{ij}}{R_{ij}^{6}} + \frac{Q_i Q_j}{\varepsilon R_{ij}} \quad (10)$$

In this potential function, Ecomplexstands for the total energy of each covalently bonded atom pairs; K_θ and K_a are the force constant for the bond angle and bond length respectively [82]; θ_{eq} and a_{eq} are the bond angle and bond length at equilibrium state [82]; β is the dihedrals and V_n is the force constant for the dihedral angles, with the γ value ranging from 0° to 180°. For the potential energy of the non-bonded atoms, which is the fourth term, consists of Van der Waals and electrostatic interactions. [82] In here, A_{ij} and B_{ij} stands for the Van der Waals terms and London dispersion terms respectively. [81] Q_i and Q_j stands for the partially charged atoms [82]; ε is the dielectric constant of the chosen medium with a value of 1 in general. [82]

To access the binding energy profile of the ligand-protein complex, a method named molecular mechanics generalized Born surface area (MM/GBSA) is applied. In MM/GBSA analysis, in addition to the molecular mechanic terms described in eq.9, the solvation energy consisting of the surface area (SA) nonpolar solvation term, the generalized Born (GB) polar solvation term and the entropy are also considered. [83] The MM/GBSA analysis can also be used to decompose the energy profile caused by a single

residue or a residue pair. [83] The binding free energy for the protein-ligand complex is estimated by eq.11:

$$\begin{aligned} E_{bind} &= E_{complex} - E_{protein} - E_{ligand} \\ &= E_{complex+} E_{GB} + E_{GA} \\ &= E_{vdw} + E_{ele} + E_{GB} + E_{GA} \end{aligned} \quad (11)$$

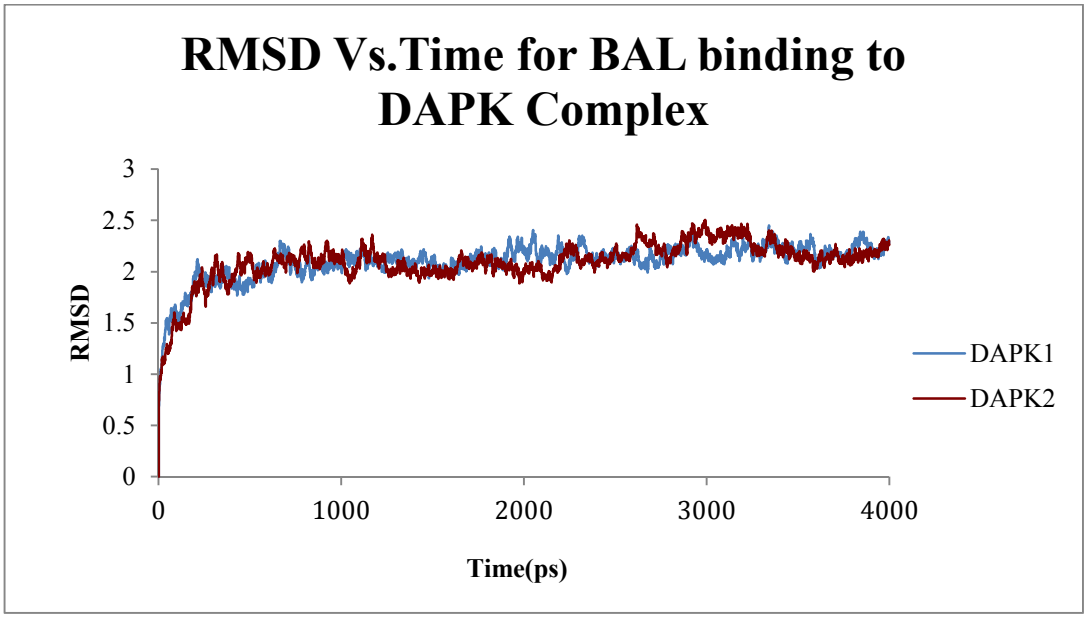
where E_{GA} and E_{GB} stands for the free energy caused by polar and nonpolar solvation respectively. [83]

The classical molecular dynamics simulation was performed to sample the conformation for the MM/GBSA analysis. Four systems were built from the docking result: mDAPK2-BAL, mDAPK2-DA7, hDAPK1-BAL, and hDAPK1-DA7. Partial charges were assigned to BAL and DA7 using Gaussian 09 (G09). Then these mol2 files containing partial charges were used to generate the topology and parameter files of the four ligand-receptor complexes via AMBER14, with the ff03 force field, followed by the 4-stage minimization. During minimization, for each stage 4000 steps of minimization and 1500 of steepest decent was applied. Then the minimized complexes were slowly heated up from 10K to 300K, over 10ps in a total of 30 stages, during the heat-up process. Next, a 16-stage production was performed at 300K over 250,000 steps. Finally, via the output files from the 16-stage MD production, the RMSD for the pair-wise atoms in the four ligand-receptor complexes were calculated and plotted against time step using Excel. During the MD simulation, the temperature and the pressure were regulated by the protocol of Andersen temperature coupling [84] and the method of isotropic position in described in David A. et al respectively [85]. Hydrogen atoms were constrained by SHAKE, with the time step was set to 2.0 fs. [86] [87] In minimization and MD production,

particle mesh Ewald (PME) was applied to manipulate the long-range electrostatic interactions. [88] The whole system was solvated in a rectangular box of TIP3P water molecules, with the dielectric constant and the cutoff for non-bonded interactions set to 1.0 and 8Å respectively. [86] The backbone (heavy atom for ligands) root mean square deviation (RMSD) is plotted against time for each snapshot in the trajectory (Figure Q.). The RMSD-time graphs show that all two trajectories reach equilibrium after 2ns.

MM/GBSA binding energy and energy decomposition analysis were performed on all four complex systems. Single snapshot protocol was used [86] to reduce the noise of energy estimation. The solvation effect was considered using the generalized Born (igb=2) method and surface area method (0.0072*ΔSASA). The last 1ns snapshots were used for binding energy and decomposition analysis to ensure the snapshots from equilibrium were used. For the simplified version of the MD simulation method, please refer to the flow chart in Sup. Figure J. 3.

(1)



(2)

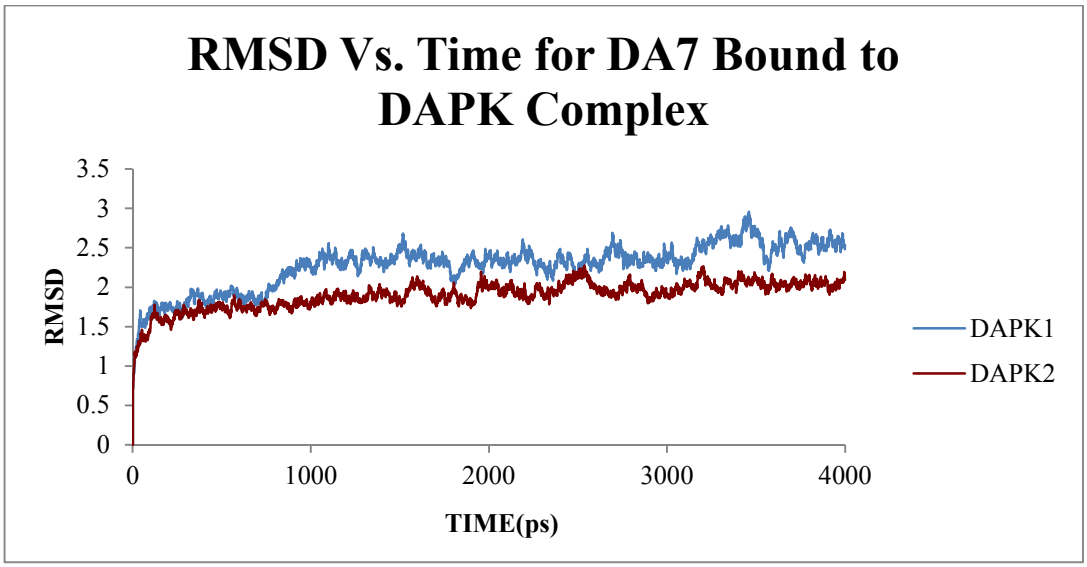


Figure S. 1-2. shows the RMSD vs. Time plot for each selected ligand bound to DAPK1 and 2 respectively X1: RMSD vs. Time for BAL bound to DAPK1 and 2 respectively. X2: RMSD vs. Time for DA7 bound to DAPK1 and 2 respectively.

Table 2 shows the binding energy of BAL and DA7 binding to hDAPK1. As the binding affinity becomes larger, the MMGBSA increases accordingly. The data listed in Table 2 indicated that BAL is a strong binder for hDAPK1.

Table 2. shows the energy profile for BAL and DA7 binding to hDAPK1. The column shows the name of the ligand –DAPK1 complex while the row shows the binding affinity of each ligand from Vina and MMGBSA energy from Amber 14 respectively.

Complex	Binding Affinity (Kcal/mol)	MMGBSA (Kcal/mol)
DAPK1-DA7	-8.9	-22.4
DAPK1-BAL	-12.3	-29.5

Table 3 below shows the binding energy of BAL and DA7 binding to mDAPK2. As the binding affinity becomes larger, the MMGBSA increases accordingly. The data listed in Table 3 indicated that DA7 is a strong binder for mDAPK2.

Table 3. shows the energy profile for BAL and DA7 binding to mDAPK2. The column shows the name of the ligand –DAPK2 complex while the row shows the binding affinity of each ligand from Vina and MMGBSA energy from Amber 14 respectively.

Complex	Binding Affinity (Kcal/mol)	MMGBSA (Kcal/mol)
DAPK2-BAL	-9.9	-26.0
DAPK2-DA7	-11.7	-47.5

MM/GBSA energy decomposition analysis is performed to figure out which residues at the interface have the largest contribution for binding specificity difference between DAPK1 and 2, the energy decomposition per-residue were performed on the four ligand complexes.

Table 4. a-b. shows the residues that contribute most to the binding specificity difference between DAPK1 and DAPK2 (cutoff >0.3kcal/mol), when ligands were bound. The sum of the energy difference for all of the interface residues between DAPK1 and 2, described in Table 4. a-b., is -2.3Kcal/mol for BAL and 3.4Kcal/mol for DA7. (See Figure P. and Sup Table 2. a-e., Sup. Table 3. a-e., Sup. Table 4. and Sup. Table 5.).

Table 4. a. Residues that contributes most to the binding specificity difference between DAPK1 and DAPK2 for BAL, cont.

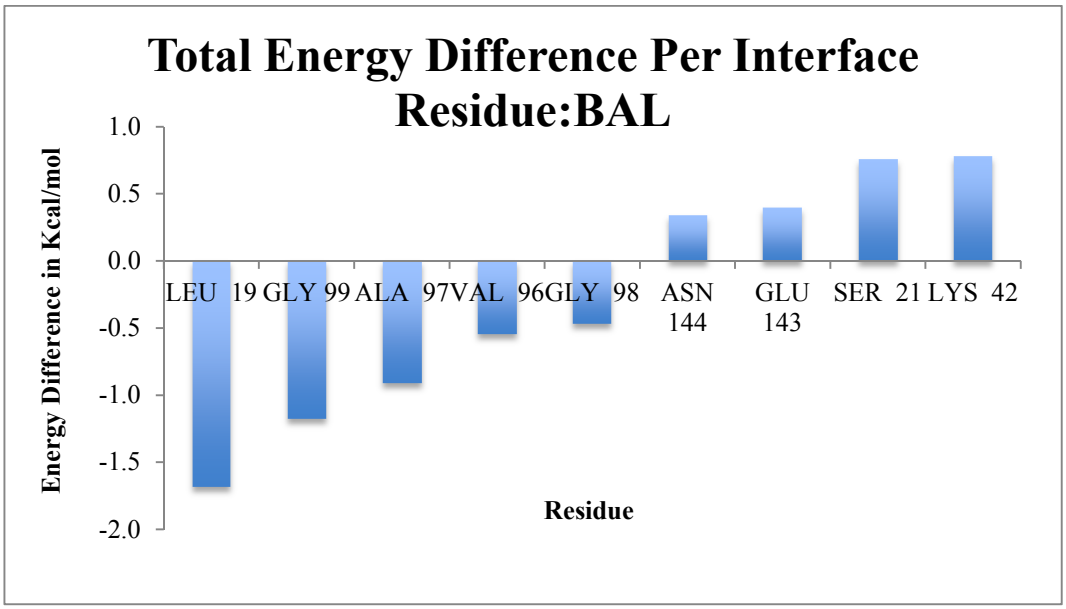
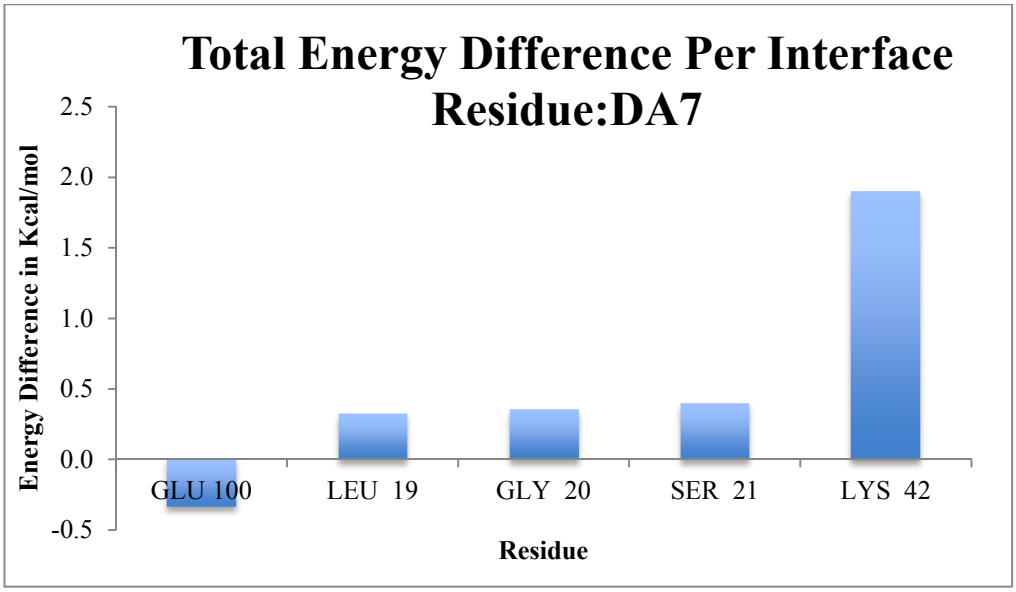


Table 4. b. Residues that contributes most to the binding specificity difference between DAPK1 and DAPK2, for DA7.



In Table 4. a., when BAL was bound to both receptors, Leu 19 and Lys42 has the largest absolute value for the energy change. In Table 4. b., when DA7 was bound to both receptors, Lys42 has the largest absolute value of the binding energy difference. To

investigate how the total energy difference at the interface residue can be reflected structurally, Figure R. 1-4. was plotted.

Figure T. 1-4. shows how the interface residues listed in Table 4. a-c. interacts with BAL and DA7 respectively. In Figure T. 1-2., Leu19, Val96, and Lys42 interact hydrophobically with BAL, as indicated by the pseudobonds colored in light pink. In Figure T. 1., looking at the interface residues, Leu19 and Lys42 are the ones whose atoms were located closely to BAL. The distance between the oxygen at the carboxyl end of Leu19 and O2 of BAL was 2.968Å. The N2 and N3 at the imidazole group of BAL also form a short distance between the side chain hydrogen of the Leu19. The distance between N2 and side chain hydrogen of HD11 and HB3 were 2.969Å and 2.412Å respectively. The distance between N3 and sidechain hydrogen HD11 and HB23 of Leu19 were 2.529Å and 2.898Å respectively. Besides Leu19, the side chain nitrogen NZ of Lys42 also interacts hydrophobically with H5 of BAL, and forms a distance of 2.283Å. In Figure T. 2., Ser 21 of DAPK1 was hydrophobically with BAL. The distance between side chain oxygen O19 and C18 and N2 imidazole group of BAL were 2.905Å and 2.892Å respectively. Based on the information provided in Figure T. 1-2. and energy decomposition results, it can be indicated that Ser21, Lys42, Leu19 and Val96 are the ones that plays the most important role in the binding specificity change when DAPK1-specific ligand was bound, due to the interaction between the ligand and the side chain atom of residues.

In Figure T. 3-4., Leu19, Gly20, Glu100, and Lys42 interacted hydrophobically with DA7, as indicated by the pseudobonds colored in light pink. In Figure T. 3., Leu19, Gly20 and Lys42 were the two interface residues interacted hydrophobically with DA7.

The distance between the residue atoms and the ligand atoms were listed in Table 5. As indicated by Table 5, DA7 encountered more constraint than BAL, from the pocket residues. In Figure T. 4., when DA7 was bound to DAPK2, Leu19 and Lys42 were the two pocket residues interacted hydrophobically with it. The distance between the side chain hydrogen HE3 on Lys 42 and C14 and C15 on DA7 is 2.497Å and 2.538Å respectively. The side chain oxygen HE3 also interacts with H11 and H12 on DA7, with a distance of 3.525Å and 3.614Å respectively. Besides Lys42, the side chain hydrogen of Leu19 also interacts with the hydrogen H7 on DA7, with a distance of 3.088Å. Based on the information provided in Figure T. 3-4. and energy decomposition results, it can be indicated that Leu19, Lys42, Glu100 and Gly20 are the ones that plays the most important role in the binding specificity change when DAPK2-specific ligand was bound, due to the interaction between the ligand and the side chain atom of residues.

Table 5. shows the closest distances between the residues atoms on Leu19, Gly20 and Lys42 and the ligand atoms, with a cutoff < 3.000Å.

Residue Atom	Ligand Atom	Distance in Å
Leu19 HD 11	DA7 C28	2.853
Leu19 HD11	DA7 C29	2.606
Leu19 HD23	DA7 C30	2.415
Leu19 HD23	DA7 C25	2.71
Leu19 HB3	DA7 H13	2.027
Leu19 O	DA7 H15	2.611
Leu19 O	DA7 H13	2.797
Leu19 O	DA7 H11	2.367
Leu19 C	DA7 H11	2.645
Gly20 N	DA7 H11	2.885
Gly20 CA	DA7 H11	2.899
GLy20 HA3	DA7 H11	2.121
GLy20 HA3	DA7 H12	2.043
GLy20 HA3	DA7 C14	2.523
GLy20 HA3	DA7 C15	2.465
Lys42 HZ2	DA7 H5	2.184
Lys42 CE	DA7 H5	2.762
Lys42 NZ	DA7 H5	2.833

To look at the complete tables for the energy decomposition data and graphs for depiction of ligands at the active pockets of DAPK1 and 2 respectively, please refer to the Sup. Table 5 and 6.

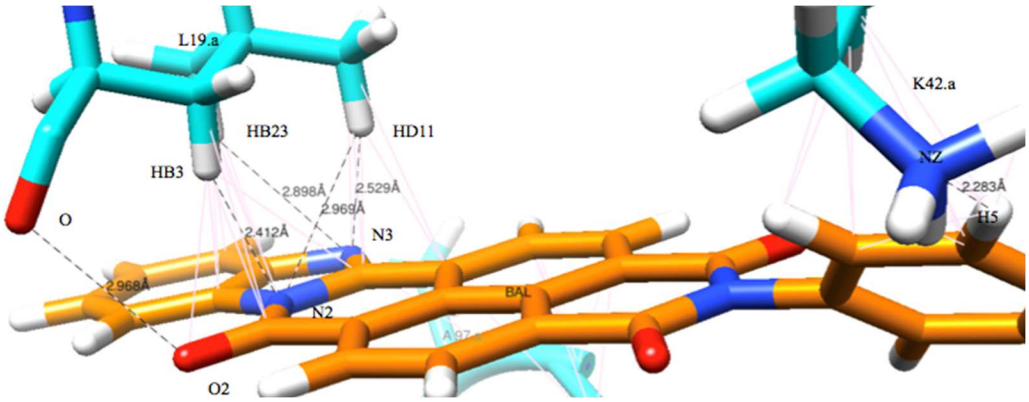


Figure T. 1. shows the conformation of the ligand bound to DAPK kinase in MD simulation, which was DAPK1-BAL, cont.

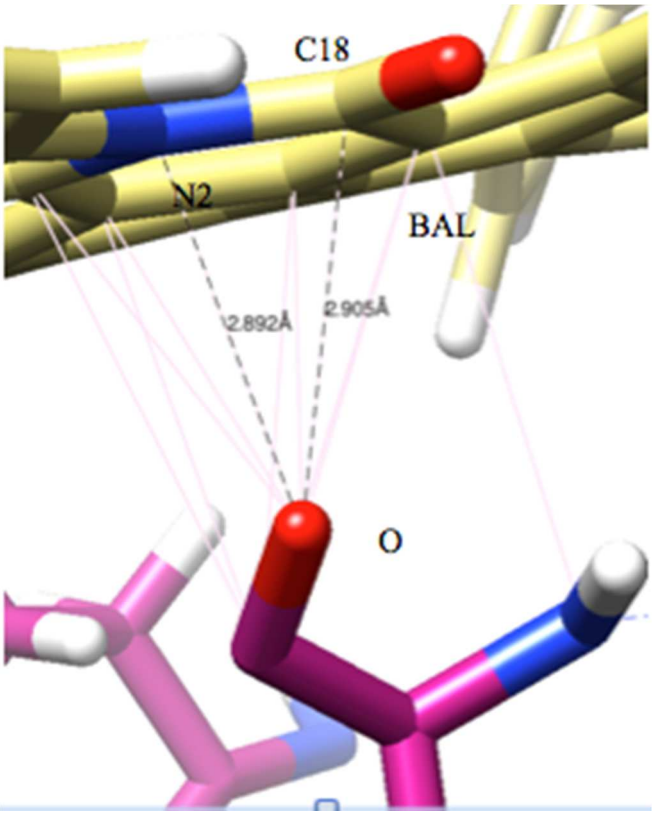


Figure T. 2. shows the conformation of the ligand bound to DAPK kinase in MD simulation, which was DAPK2-BAL, cont.

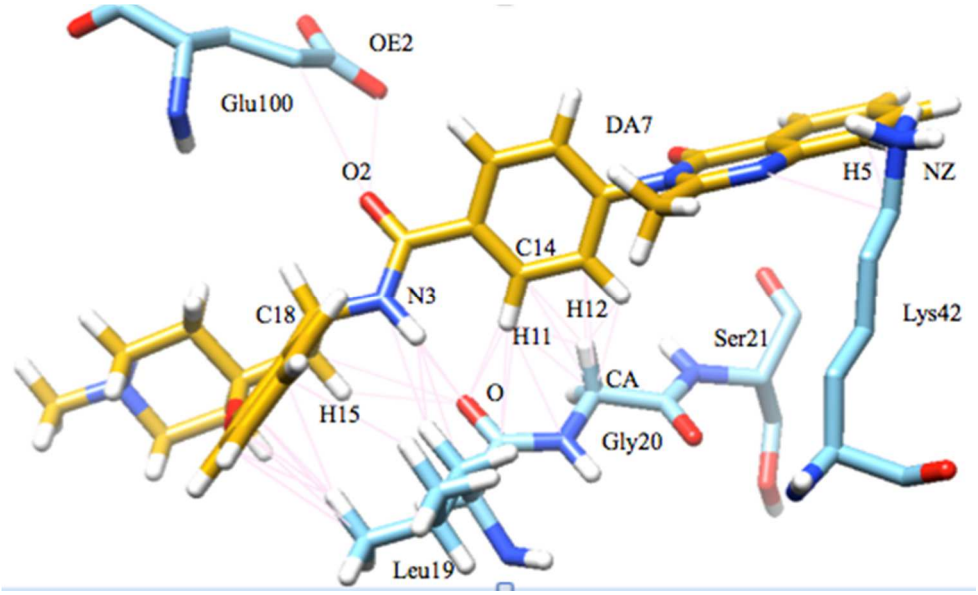


Figure T. 3. shows the conformation of the ligand bound to DAPK kinase in MD simulation, which was DAPK1-DA7, cont.

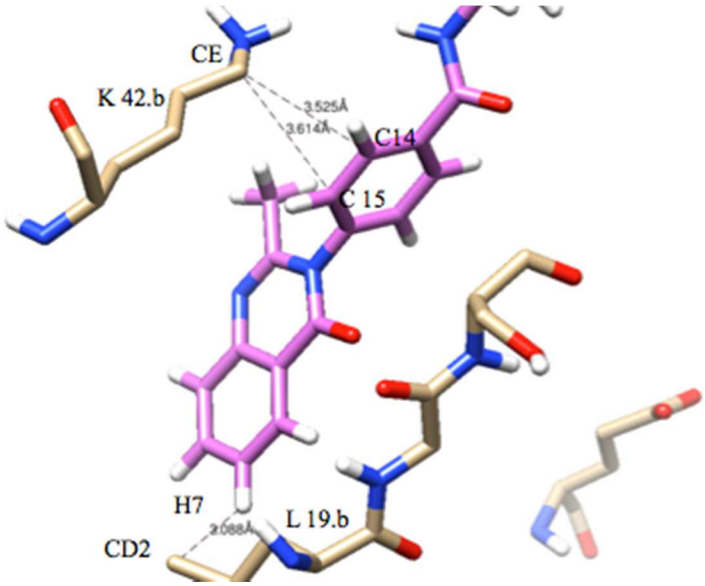


Figure T. 4. shows the conformation of the ligand bound to DAPK kinase in MD simulation, which was DAPK2-DA7, cont.

III. Docking on ELMO1-DOCK2 Complex

1.Preparation before Virtual Screening

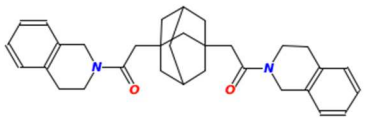
AutoDock Vina, software designed specifically for virtual screening, is selected to perform file conversion and predict the binding affinity during virtual screening. All of the ligands used in both rigid and flexible docking processes were downloaded from ZINC database [4][5]. The level of exhaustiveness was set to 1 for rigid docking process while for flexible docking process it was set to 9. In both rigid docking and flexible docking, the number of model was set to 1 and energy range was set to 3.

2. Docking Inhibitors on ELMO1-DOCK2 Complex at SH3 Pocket

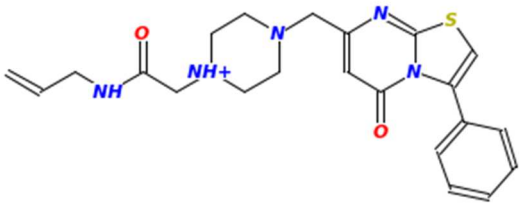
To get an overview about the types of ligands that have higher binding affinity with the SH3 pocket on ELMO1-DOCK2 complex, a subset called Clean Fragment-Like Molecules (~16M) was tested in the rigid docking process. The structure of the receptor was taken from the PDB ID 3A98 and converted from pdb into pdbqt format by script prepare_receptor4.py in Autodock Vina. [52] [7] Then the x, y and z value of center coordinates and the magnitudes of length, width and height of a docking cube are determined from the coordinates of the pocket residues of located at the SH3 Pocket. When the docking process finished, the results from the rigid docking process were top ~1M ligand candidates were selected to run the flexible docking. See Sup. Figure Q. for the flow chart of the docking method. The steps for the flexible docking process were similar to the flexible docking process. See Sup. Figure R. for the flow chart of the docking method.

Figure U. 1-10. shows the top ten ligand-candidates among the top 1M molecules

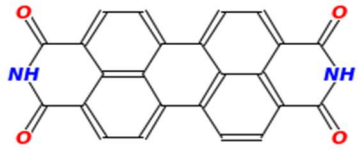
obtained from the previous rigid docking process at SH3 pocket, in flexible docking process (Exhaustiveness = 9). These ligand candidates were ranked solely by the binding affinity. Compared with the results from the rigid docking process (Sup. Figure T.), only the ligand at the 10^{th} place was different. Based on the screening results on SH3 domain, it can be indicated that SH3 might not be an ideal pocket for inhibitors to disrupt ELMO1-DOCK2 complexes.



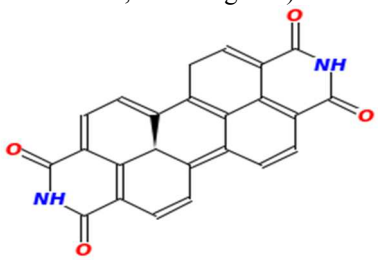
1. ZINC22013692
(-8.8Kcal/mol, 442.478g/mol)



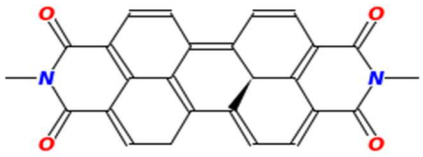
2. ZINC22013688
(-8.6Kcal/mol, 442.478g/mol)



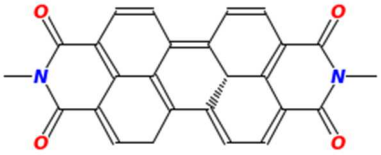
3. ZINC00640220
(-8.6Kcal/mol, 390.354g/mol)



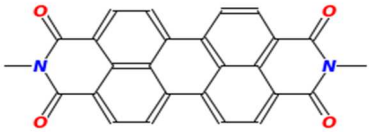
4. ZINC71257150
(-8.5Kcal/mol, 392.37g/mol)



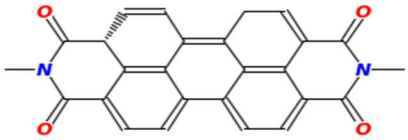
5. ZINC71257153
(-8.4Kcal/mol, 420.424g/mol)



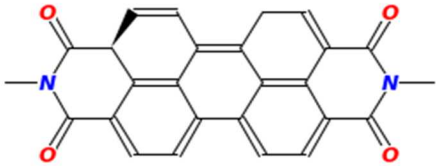
6. ZINC71257152
(-8.4Kcal/mol, 420.424g/mol)



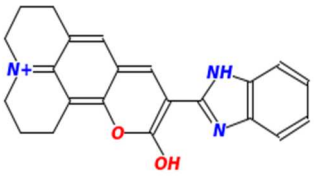
7. ZINC11612116
(-8.3Kcal/mol, 418.408g/mol)



8. ZINC22003706
(-8.3Kcal/mol, 420.424g/mol)



9. ZINC22003711
(-8.2Kcal/mol, 420.424g/mol)



10. ZINC37865607
(-8.1Kcal/mol, 422.552g/mol)

Figure U. 1-10. Top 10 Ligand Candidates for ELMO1-DOCK2 complex at SH3 pocket in flexible docking, with their corresponding ZINC ID (exhaustiveness = 9). These ligands were numbered based on their binding affinity.

3. Preliminary Screening on ELMO1-DOCK2 Complex at the Helix Pocket

The Clean Drug-Like subset with Tanimoto coefficient = 0.9 down loaded from ZINC [4][5] was used to perform the initial screening on the helix site flexibly. The steps for the flexible docking process were similar to the ones mentioned in Step **2**. Finally, the top 1K molecules with highest binding affinity were selected to perform the ligand-based screening on the ~13M Clear-Fragment Like Drugs fragment mentioned in Step **4**. See Sup. Figure S. for the flow chart of the docking method. For the initial screening results at the helix pocket, please refer to Sup. Figure V.

4. Ligand-Based Screening on the Top 1K Candidates from ~12.3K Clean Drug-Like Molecules

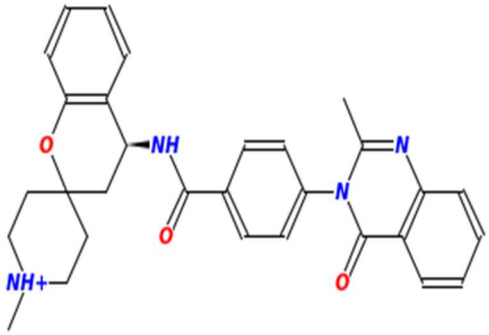
Initially, the original *.smi file containing ~13M Clean Drug-Like molecules (Tanimoto coefficient level = 100%) were converted to the fast search index format (*.fs). Then the top 1K molecules with highest binding affinity, from Step **3**, was selected to compared with the ~13M molecules saved in the fast-search index file. In this process, after removing the duplicate ligand candidates, ~480,202 molecules with Tanimoto coefficient >0.6 were extracted and used to perform the next docking steps. See Sup. Figure T. for the flow chart of the docking method.

5. Secondary Screening on ELMO1-DOCK2 Complex at the Helix Pocket

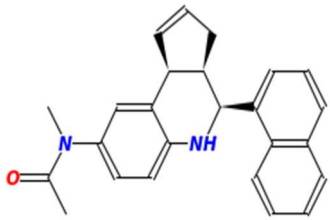
To search for more ideal ligand candidates for ELMO1-DOCK2 complex, a secondary screening was performed on the ~480,202 molecules obtained in Step **4**. The methods used for this test was similar to the ones described in Step **2**. In the end, the binding affinity of these ligands were analyzed and ranked, with their corresponding ZINC ID extracted. See Sup. Figure U. for the flow chart of the docking method. Figure

V. 1-10. and Sup. Figure W. shows the top ten ligand-candidates for ELMO1-DOCK2 complex at helix pocket during secondary screening, in flexible docking (Exhaustiveness = 9). These ligand candidates were ranked solely by the binding affinity, among the ~480,202 Clean-Fragment Like molecules from Step **4.**

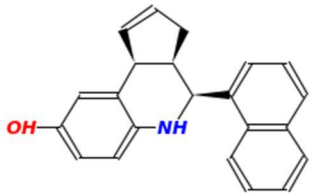
Based on the screening results that the structures of ligand candidates weren't so similar as the ones binding at the SH3 pocket, a conclusion can be drawn that the interface pocket might locate at the six-bundle helix formed by ELMO1-DOCK2.



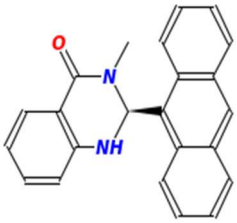
1. ZINC48368120
(-8.3Kcal/mol, 482.668g/mol)



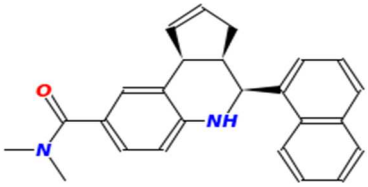
2. ZINC20698566
(-8.2Kcal/mol, 368.48g/mol)



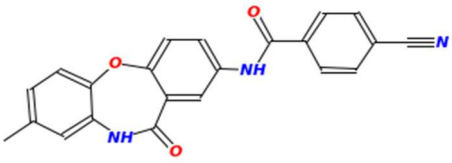
3. ZINC14463638
(-8.1Kcal/mol, 313.4g/mol)



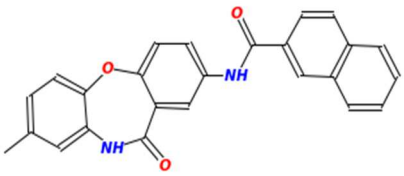
4. ZINC66128373
(-8.1Kcal/mol, 338.41g/mol)



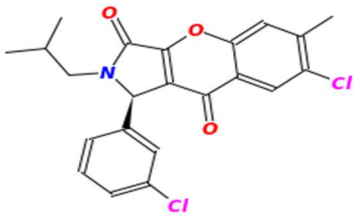
5. ZINC20699276
(-8.1Kcal/mol, 368.48g/mol)



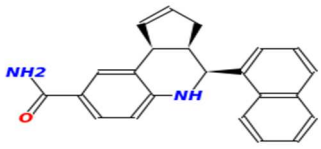
6. ZINC38715363
(-8.0Kcal/mol 369.38g/mol)



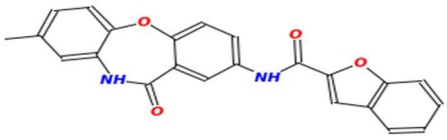
7. ZINC38715402
(-8.0Kcal/mol, 394.43g/mol)



8. ZINC38715378
(-8.0Kcal/mol, 380.35g/mol)



9. ZINC3871538010
(-8.0Kcal/mol, 340.426g/mol)



10. ZINC14463744
(-8.0Kcal/mol, 384.391g/mol)

Figure V. 1-10. shows the top ten ligand-candidates for ELMO1-DOCK2 complex at helix pocket during secondary ligand-based screening, in flexible docking (Exhaustiveness = 9). These ligand candidates were ranked solely by the binding affinity, among the ~480,202 Clean-Fragment Like molecules from Step **4**.

Chapter 3 Discussion

I. Factors Affecting Binding Specificity of DAPK2 and Future Direction

Structural-based virtual screening was carried out to identify inhibitors for DAPK2 specifically using AutoDock Vina, followed by the molecular dynamic simulation to test the ligand specificity and MM/GBSA analysis per residue by AMBER. Our study shows that based the binding specificity of the DAPK kinase was affected by the distance of the atom pairs located on the designated ligand candidate and its nearby residues respectively. Leu19, Gly20, Val27, Ser21, Lys42 and Glu100 are found to be the residues that contribute most to the binding specificity of DAPK kinase. In Figure.T1 and T2, since BAL has more hydrophobic contacts in DAPK1 than in DAPK2, it can be concluded that BAL is more specific towards DAPK1. In Figure.T3 and T4, although the energy decomposition result shows that DA7 has higher binding affinity and MMGBSA, DA7 has more hydrophobic constraint in DAPK1. Depicted from the docking and MD simulation results, DA7 might be the most competitive ligand for DAPK2, although the molecular weight isn't quite ideal for drug design. Since the binding pocket for DAPK2 was designed based on the binding pocket residues for DAPK1, the selected ligand candidates might also have more hydrophobic contacts towards the DAPK1 pocket residues.

Since DAPK1 and 2 were structurally similar to each other, sensitivity is an issue that should be pay attention. If necessary, a secondary screening needs conducting, with the first molecule listed in Figure U. as the reference structure. Besides conducting a

secondary screening, structural mutation can also be performed on these two DAPK kinases, by changing the position of the residue into the one it supposed to have in another structure. In this way, by using a smaller ligand library, the results obtained from the previous screening process can be examined to see if they were the true ligands for DAPK2 with high specificity. Finally, the virtual screening method nowadays depends deeply on the class of compound activities. To reduce the side effects caused by the rejection of candidates with a certain method, different methods should also be applied at the same time during the consensus scoring schemes.

II. Factors Affecting Inhibitors Bindng to ELMO1-DOCK2 Complex and Future Direction

Structural-based virtual screening was carried out to identify inhibitors for ELMO1－DOCK2 complex specifically via structural-based virtual screening. Our study shows several residues were related to the binding interactions between the ligands and the ELMO1-DOCK2 complex at the SH3 domain and five-bundle helices as well.

At the pocket between SH3 binding domain and the Pxxp domain, the hydrogen at hydroxyl side chain of Ser 22 form a hydrogen bond with the O1 at carbonyl group of ZINC22013692 displayed in Figure W. The hydrogen bond distance is also labeled in Figure W, which is measured to be 3.534Å. In addition, several pseudo bonds were also detected between the ligand and the residues at or near the binding pockets, which were ser22, Pro711, Pro712, ILE 713, and Pro 714 at the Pxxp domain.

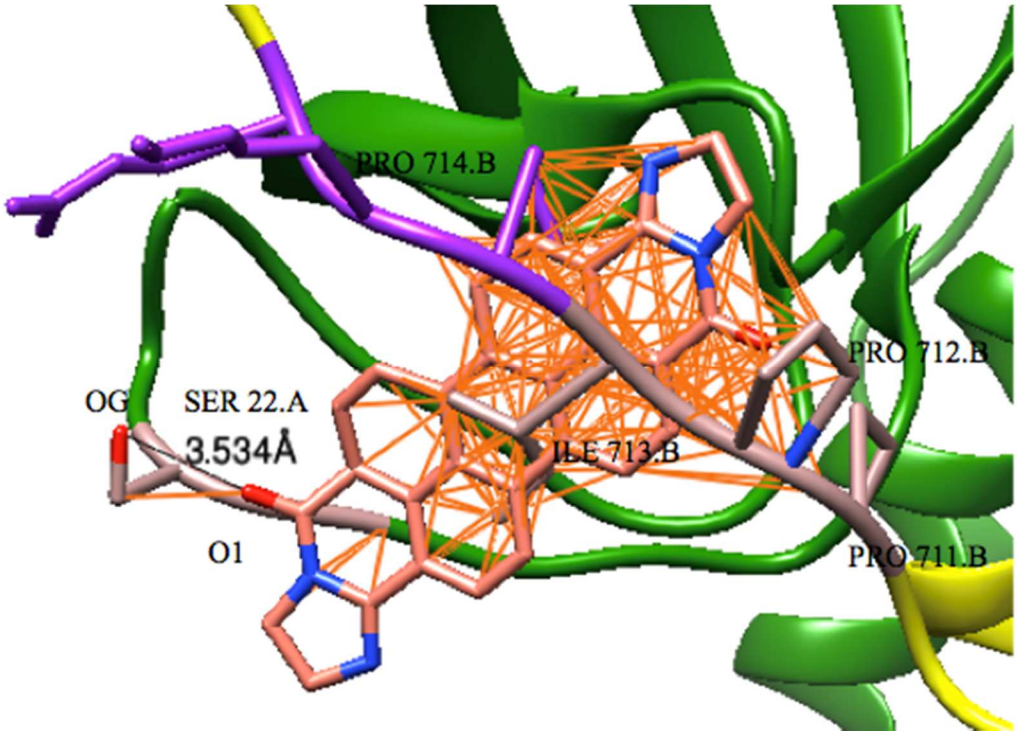


Figure W. shows the interactions occurred between ZINC22013692 and its nearby residues. Hydrogen bond between the hydrogen at hydroxyl side chain of Ser22 (in rosy brown) and the O1 at carbonyl group of ligand 1 (in pink) was colored in black, with the distance displayed. In addition, pseudo-bonds (in orange) were also detected between the ligand and the residues near or at the binding pocket, coloring in rosy pink as well.

At the five-bundle helices pocket, the hydrogen attached to the nitrogen of indole sidechain of Trp96 forms a hydrogen bond with the carbonyl oxygen of ZINC48368120 displayed in Figure X. The hydrogen bond distance is also labeled in Figure I., which is measured to be 3.174Å. In Figure Y., several pseudo bonds were also detected between the ligand and the residues at or near the binding pocket, which were Trp96, Glu693, Leu696 and Leu699.

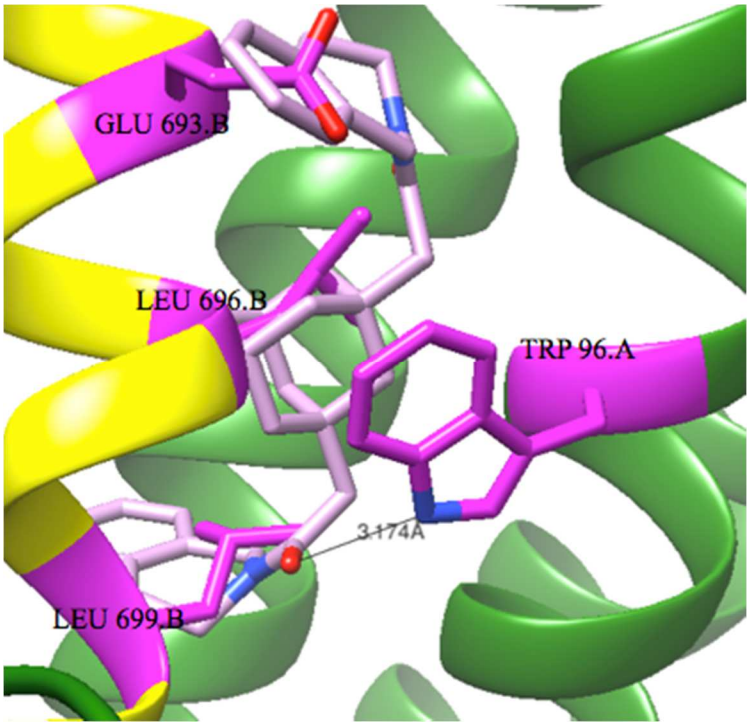


Figure X. shows the interactions occurred between the ZINC48368120 and its nearby residues. Hydrogen bond between the hydrogen attached to the indole nitrogen of Trp96 (in dark magenta) and the O2 at carbonyl group of ZINC48368120 (in pink) was colored in black, with the distance displayed.

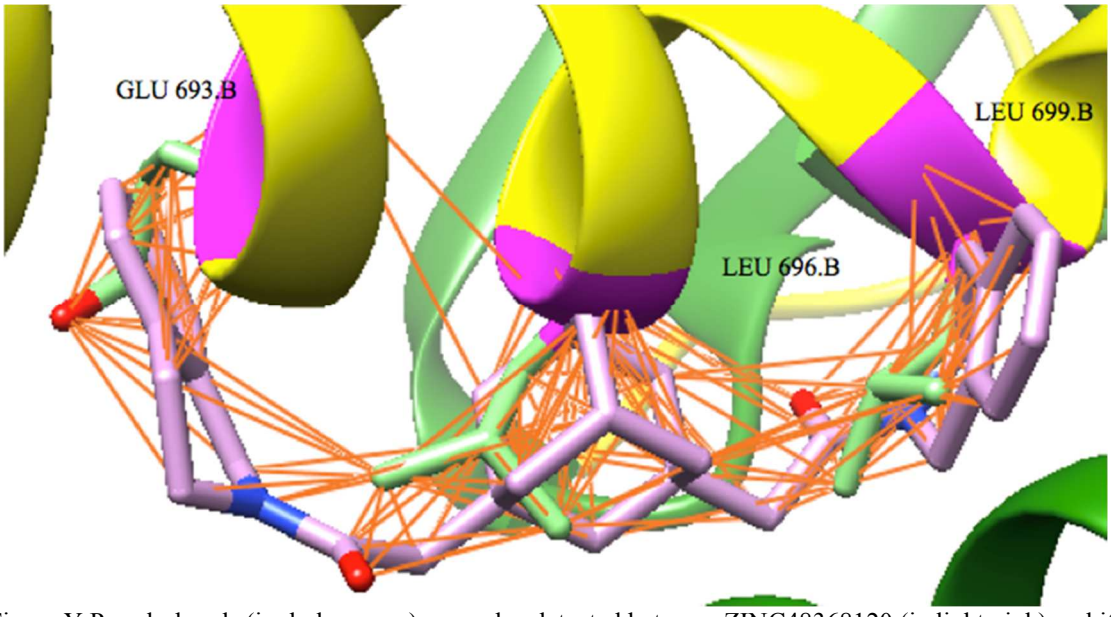


Figure Y Pseudo-bonds (in dark orange) were also detected between ZINC48368120 (in light pink) and its nearby residues (backbone in magenta and sidechain in light green), which were Leu699, Leu696, and Glu693.

During the process of fingerprint screening, the top 1,0000 ligands from the Clean Drug-Like subset were compared with the Clear-Fragment Like Drugs. Since the number of reference molecules used is limited, the fingerprint screening method is dependent heavily on how similar the ligands are to the referenced structure, regardless of their unique characteristic. As a result, some potential candidates might be rejected. To minimize the loss of attractive candidates, a consensus-scoring scheme can be applied via the usage of multiple screening methods at the same time. [24]

At the SH3 binding domain, the structural studies show that Ser22, Pro711, Pro712, ILE 713, and Pro714 at the Pxxp domain interacts with the ligand via the formation of hydrogen bonds or some other steric interactions at the side chains. According to the results from the previous research, ELMO1 bound to DOCK2 via the proline residues on the Pxxp domain. In this way, the ELMO1-DOCK2 protein complex is formed in a rigid and large fashion.

At the five-bundle helices, the result of the structural analysis shows that Glu693, Leu696, and Trp96 at the helices of ELMO1 interacts with the ligand via the formation of hydrogen bonds or some other steric interactions at the side chains. According to the results from the previous research, Trp96 and Glu694 were two of the residues at DOCK2 interact with the Leu696 and Leu699 at ELMO1. The formation of the ligand can in a way disrupts the hydrophilic interactions between ELMO1 and DOCK2 and thus prevents the formation of the complex.

For the selection of ligands, experimental binding assay will be conducted to find out the best candidates for drug design. In addition, some structural analysis will also be

carried out to determine which binding pocket is the ultimate one for the disruption of the ELMO1-DOCK2 complex.

Chapter 4 Conclusion

Death associated protein kinase 2 (DAPK2) is a serine threonine kinase with 370 amino acid residues, whose structure is highly conserved with the other members in the death associated protein kinase (DAPK) family. Although the structures of DAPK1 and 2 are highly similar to each other, currently few studies regarding the drug design of DAPK2 has been conducted. To find out the drug candidates with high specificity to DAPK2, structural-based virtual screening was carried out. Based on the structural and sequence alignment results between DAPK1 and 2, a virtual screening protocol was developed including the rigid docking and flexible docking on the identified DAPK2 binding site, the construction of ligand clusters based on their Tanimoto coefficient. Furthermore, based on their binding score in DAPK1 and 2 respectively, BAL and DA7 were the two candidates selected to analyze which residues could cause the binding specificity difference between DAPK2 and DAPK1. According to the results of docking, ligand clustering, MM/GBSA analysis and position of ligands in the DAPK2 binding pocket, it can be concluded that DA7, whose binding affinity and MM/GBSA score are -11.7 Kcal/mol and -47.5 Kcal/mol respectively, is the most ideal candidate so far for DAPK2.

Engulfment and cell motility 1 protein (ELMO1) is a human protein regulated by ELMO1 gene, with 720 residues in total. It plays an important role in clearing the apoptotic germ cells in vivo. Dedicator of cytokinesis 2 (DOCK2) is a guanine nucleotide exchange factor with a mass of 213 kDa, which is responsible for the activation of the chemotaxis process. When interacting with each other, DOCK2 and ELMO1 forms a large, rigid complex via the SH3 domain at DOCK2 or the formation of

the five-bundle helices by ELMO1 and DOCK2. To identify the potent drug candidates that interrupt the formation of the ELMO1-DOCK2 complex, both structural- and ligand-based virtual screening were carried out at the two potential binding pockets: one is located at the SH3 domain of DOCK2 and the other is located between the 5-helix bundles. From the rigid and flexible docking results, ZINC22013692 and ZINC48368120 were candidates with the highest binding affinity for the pocket at the SH3 domain and 5-helix bundles respectively. For the pocket at SH3 domain, Ser22, Pro711, Pro712, ILE 713, and Pro714 at the Pxxp domain were found to interact with the ligand via the formation of hydrogen bonds or some other steric interactions at the side chains. For the pocket at the 5-helix bundles, Glu693, Leu696, and Trp96 at the helices of ELMO1 were found to interact with the ligand and thus the ligand candidate can disrupt the hydrophilic interactions between ELMO1 and DOCK2 to prevent the formation of the complex.

These two case studies provide a general notion for the combination of structural-based screening, ligand-based screening and MM/GBSA energy analysis in drug design, which can be a milestone on the future drug design of DAPK receptors.

Reference

1. Bajorath, Jürgen. "Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications." Nature Reviews 3.11 (2004): 12. Print.

2. Schneider, G. "Virtual Screening: An Endless Staircase?" Nat Rev Drug Discov 9.4 (2010): 273-6. Print.

3. Solt, Ivan. "New Approaches to Virtual Screening." Drug Discovery and Development 16.5 (2013): 12. Print.

4. Srinivas Reddy, S. Priyadarshini Pati, P. Praveen Kumar, H.n. Pradeep, and Narahari Sastry. "Virtual Screening in Drug Discovery - A Computational Perspective." Current Protein & Peptide Science CPPS 8 (2007): 329-51. Print.

5. Vidal, D., M. Thormann, and M. Pons. "A Novel Search Engine for Virtual Screening of Very Large Databases." Journal of Chemical Information and Modeling 46.2 (2006): 836-43. Print.

6. Lemmen, C., T. Lengauer, and G. Klebe. "Flexs: A Method for Fast Flexible Ligand Superposition." J Med Chem 41.23 (1998): 4502-20. Print.

7. Schneider, G. "Virtual Screening: An Endless Staircase?" Nat Rev Drug Discov 9.4 (2010): 273-6. Print.

8. Thorner, D. A. "Similarity Searching in Files of Three-Dimensional Chemical Structures: Flexible Field-Based Searching of Molecular Electrostatic Potentials." Journal of Chemical Information and Computer Sciences 36.4 (1996): 900-08. Print.

9. Mestres, J, D C Rohrer, and G M Maggiora. “A Molecular Field-Based Similarity Approach to Pharmacophoric Pattern Recognition.” Journal of molecular graphics & modelling 15.2 (1997): 114–21, 103–6. Web. 24 Feb. 2016.

10. Krämer, Andreas, Hans W. Horn, and Julia E. Rice. “Fast 3D Molecular“Fast 3D Molecular Superposition and Similarity Search in Databases of Flexible Molecules.” *Journal of Computer-Aided Molecular Design* 17.1 (2003): 13–38. Web.

11. Todeschini, R., and M. Lasagni. "New Molecular Descriptors for 2d and 3d Structures - Theory." Journal of Chemometrics 8 4 (1994): 263-72. Print.

12. Floris, M. "A Generalizable Definition of Chemical Similarity for Read-Across." Journal of Cheminformatics 6 (2014). Print.

13. Brown, Robert D, and Yvonne C Martin. “Use of Structure-Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection.” Journal of Chemical Information and Computer Sciences 36.3 (1996): 572–584. Web.

14. Joørgensen, A. M. “A Fragment-Weighted Key-Based Similarity Measure for Use in Structural Clustering and Virtual Screening.” QSAR and Combinatorial Science 25.3 (2006): 221–234. Web.

15. Kim, Sunghwan. “PubChem Substance and Compound Databases.” Nucleic acids research 44.D1 (2015): D1202–1213. Web.

16. Butina, Darko. “Unsupervised Data Base Clustering Based on Daylight’s Fingerprint and Tanimoto Similarity: A Fast and Automated Way to Cluster Small and Large Data Sets.” Journal of Chemical Information and Computer Sciences 39.4 (1999): 747–750. Web.

17. Eckert, Hanna, and Jürgen Bajorath. “Determination and Mapping of Activity-Specific Descriptor Value Ranges for the Identification of Active Compounds.” Journal of Medicinal Chemistry 49.7 (2006): 2284–2293. Web.

18. Rarey, M, and J S Dixon. “Feature Trees: A New Molecular Similarity Measure Based on Tree Matching.” Journal of computer-aided molecular design 12.5 (1998): 471–490. Web.

19. Bocker, A. “A Hierarchical Clustering Approach for Large Compound Libraries.” Journal of Chemical Information and Modeling 45.4 (2005): 807–815. Web.

20. Oprea, Tudor I., and Hans Matter. “Integrating Virtual Screening in Lead Discovery.” Current Opinion in Chemical Biology 2004: 349–358. Web.

21. Goodford, P J. “A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules.” Journal of medicinal chemistry 28.7 (1985): 849–857. Web.

22. Mason, J S, and D L Cheney. “Ligand-Receptor 3-D Similarity Studies Using Multiple 4-Point Pharmacophores.” Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing 467 (1999): 456–67. Web.

23. Ahlstrom, M. “CYP2C9 SAR: Optimize Metabolic Stability COX-2 Inhibitor.” J. Med. Chem. 50 .18 (2007): 4444–4452. Web.

24. Haigh, James A. “Small Molecule Shape-Fingerprints.” Journal of Chemical Information and Modeling 45.3 (2005): 673–684. Web.

25. Ronkko, Toni. “BRUTUS: Optimization of a Grid-Based Similarity Function for Rigid-Body Molecular Superposition. II. Description and Characterization.” Journal of Computer-Aided Molecular Design 20.4 (2006): 227–236. Web.

26. Khan, Feroz. “Modern Methods & Web Resources in Drug Design & Discovery.” Letters in Drug Design & Discovery 8.5 (2011): 469–490. Web.

27. Taylor, R. D., P. J. Jewsbury, and J. W. Essex. “A Review of Protein-Small Molecule Docking Methods.” Journal of Computer-Aided Molecular Design 16.3 (2002): 151–166. Web.

28. Metropolis, Nicholas. “Equation of State Calculations by Fast Computing Machines.” The Journal of Chemical Physics 21.6 (1953): 1087–1092. Web.

29. Yang, Yian, Jiaqiang Qian, and Dengming Ming. “Docking Polysaccharide to Proteins That Have a Tryptophan Box in the Binding Pocket.” Carbohydrate Research 414 (2015): 78–84. Web.

30. Žiemys, L. Rimkute, J. Kulys. "Modeling the Enantioselective Enzymatic Reaction with Modified Genetic Docking Algorithm." Nonlinear Analysis: Modelling and Contro 9.4 (2004): 10. Print.

31. Morris, Garrett M. “Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function.” Journal of Computational Chemistry 19.14 (1998): 1639–1662. Web.
32. Lemmen, C., T. Lengauer, and G. Klebe. "Flexs: A Method for Fast Flexible Ligand Superposition." J Med Chem 41.23 (1998): 4502-20. Print.

33. Welch, W, J Ruppert, and A N Jain. “Hammerhead: Fast, Fully Automated Docking of Flexible Ligands to Protein Binding Sites.” Chemistry & Biology 3.6 (1996): 449–462. Web.

34. Huang, Sheng-You, Sam Z Grinter, and Xiaoqin Zou. “Scoring Functions and Their Evaluation Methods for Protein-Ligand Docking: Recent Advances and Future Directions.” Physical chemistry chemical physics: PCCP 12.40 (2010): 12899–12908. Web.

35. Chang, M. W. "Virtual Screening for Hiv Protease Inhibitors: A Comparison of Autodock 4 and Vina." PLoS One 5.8 (2010): e11955. Print.

36. Trott, Oleg, and Arthur J. Olson. "AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading." J. Comput. Chem. Journal of Computational Chemistry 31.2 (2010): 455-61. Print.

37. Skajaa, Anders. Limited Memory BFGS for Nonsmooth Optimization. Thesis. New York University, 2010. N.p.: n.p., n.d. Print.

38. Talbi, El-Ghazali. "On the Efficiency of Local Search Methods for the Molecular Docking Pr." Oblem. Springer-Verlag Berlin Heidelberg, 2009. Web. 14 Oct. 2015.

39. Abagyan, Ruben, Maxim Totrov, and Dmitry Kuznetsov. "ICM? A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation." J. Comput. Chem. Journal of Computational Chemistry 15.5 (1994): 488-506. Web.

40. Maigret, Bernard, Ghemtio, and Jeanot. "Efficiency of a Hierarchical Protocol for High Throughput Structure-based Virtual Screening on GRID5000 Cluster Grid." OAB Open Access Bioinformatics (2010): 41. Web.

41. Patel, Ashok K., Ravi P. Yadav, Viivi Majava, Inari Kursula, and Petri Kursula. "Structure of the Dimeric Autoinhibited Conformation of DAPK2, a Pro-Apoptotic Protein Kinase." Journal of Molecular Biology 408.3 (2011): 369-83. Print.

42. Shiloh, R., S. Bialik, and A. Kimchi. "The Dapk Family: A Structure-Function Analysis." Apoptosis 19 2 (2014): 286-97. Print.

43. Okamoto, Masako, Kiyoshi Takayama, Tomoko Shimizu, Kazuhiro Ishida, Osamu Takahashi, and Toshio Furuya. "Identification of Death-Associated Protein Kinases Inhibitors Using Structure-Based Virtual Screening." J. Med. Chem. Journal of Medicinal Chemistry 52.22 (2009): 7323-327. Print.

44. Wilbek, Theis S., Tine Skovgaard, Fiona J. Sorrell, Stefan Knapp, Jens Berthelsen, and Kristian Strømgaard. "Identification and Characterization of a Small-Molecule Inhibitor of Death-Associated Protein Kinase 1." ChemBioChem 16.1 (2014): 59-63. Print.

45. Carlessi, R., Levin-salomon, V., Ciprut, S., Bialik, S., Berissi, H., Albeck, S., … Kimchi, A. (2011). Scientific report. Nature Publishing Group, 12(9), 917–923. http://doi.org/10.1038/embor.2011.126

46. McNamara, Laurie K. “Site-Directed Mutagenesis of the Glycine Rich Loop of Death Associated Protein Kinase (DAPK) Identifies It as a Key Structure for Catalytic Activity.” *Biochimica et biophysica acta* 210518 (2010): n. pag. Web.

47. Velentza, A.V. “An Aminopyridazine-Based Inhibitor of a pro-Apoptotic Protein Kinase Attenuates Hypoxia-Ischemia Induced Acute Brain Injury.” Bioorg.Med.Chem.Lett. 13 (2003): 3465–3470. Web. 19 Feb. 2016.

48. Ueda, Y. “Complex Structure of Kinase Domain of DAP Kinase with BDB402.” To be Published n. pag. Web. 19 Feb. 2016.

49. Feng, L. “Structurally Sophisticated Octahedral Metal Complexes as Highly Selective Protein Kinase Inhibitors.” J.Am.Chem.Soc. 133 (2011): 5976. Web. 19 Feb. 2016.

50. Kursula, P, and M. Wilmanns. “Crystal Structure of Human Drp-1 Complexed with an Inhibitor.” To be Published n. pag. Web. 23 Feb. 2016.

51. Yokoyama, T, Y Kosaka, and M. Mizuguchi. “Structural Insight into the Interactions between Death-Associated Protein Kinase 1 and Natural Flavonoids.” J.Med.Chem. 58 (2015): 7400–7408. Web. 20 Feb. 2016.

52. Pike, A.C.W. “Activation Segment Dimerization: A Mechanism for Kinase Autophosphorylation of Non-Consensus Sites.” Embo J. 27 (2008): 704. Web. 19 Feb. 2016.

53. Huber, K. “7, 8-Dichloro-1-Oxo-Beta-Carbolines as a Versatile Scaffold for the Development of Potent and Selective Kinase Inhibitors with Unusual Binding Modes.” J.Med.Chem. 55 (2012): 403–413. Web. 19 Feb. 2016.

54. Filippakopoulos, P. “Crystal Structure of Human Death Associated Protein Kinase 3 (DAPK3) in Complex with an Imidazo-Pyridazine Ligand.” To be Published n. pag. Web. 19 Feb. 2016.

55. Rodrigues. “De Novo Fragment Design for Drug Discovery and Chemical Biology.” Angew.Chem.Int.Ed.Engl. 54 (2015): 15079. Web. 19 Feb. 2016.

56. Schumacher, A. M. "Death Associated Protein Kinase (Dapk) as a Potential TherapeuticTarget in Brain Injury: Measurement of Quantitative Enzyme Activity Changes in an Animal Model and Development of the First Small Molecule Inhibitor with in Vivo Activity." Faseb Journal 17.5 (2003): A1033-A33. Print.

57. Shamloo, M. "Death-Associated Protein Kinase Is Activated by Dephosphorylation in Response to Cerebral Ischemia." Journal of Biological Chemistry 280.51 (2005): 42290-99. Print.

58. Pelled, D. "Death-Associated Protein (Dap) Kinase Plays a Central Role in Ceramide-Induced Apoptosis in Cultured Hippocampal Neurons." Journal of Biological Chemistry 277.3 (2002): 1957-61. Print.

59. Yamamoto, M. "Dap Kinase Activity Is Critical for C-2-Ceramide-Induced Apoptosis in Pc12 Cells." European Journal of Biochemistry 269.1 (2002): 139-47. Print.

60. Komander, David. “An A-Helical Extension of the ELMO1 Pleckstrin Homology Domain Mediates Direct Interaction to DOCK180 and Is Critical in Rac Signaling.” Ed. Josephine C. Adams. Molecular Biology of the Cell 19.11 (2008): 4837–4851. PMC. Web. 18 Aug. 2015.

61. Hanawa-Suetsugu, Kyoko. "Structural Basis for Mutual Relief of the Rac Guanine Nucleotide Exchange Factor DOCK2 and Its Partner ELMO1 from their Autoinhibited forms." National Center for Biotechnology Information. U.S. National Library of Medicine, 28 Feb. 2012. Web. 18 Aug. 2015.

62. Sanui, Terukazu. "DOCK2 Regulates Rac Activation and Cytoskeletal Reorganization through Interaction with ELMO1." National Center for Biotechnology Information. U.S. National Library of Medicine, 15 Oct. 2003. Web. 18 Aug. 2015.

63. Kunisaki, Yuya. "Dock2 Is a Rac Activator That Regulates Motility and Polarity During Neutrophil Chemotaxis." The Journal of Cell Biology 174 5 (2006): 6. Print.

64. Namekata, K., Kimura, A., Kawamura, K., Harada, C., & Harada, T. (2014). Dock GEFs and their therapeutic potential: Neuroprotection and axon regeneration. Progress in Retinal and Eye Research, 43, 1–16. http://doi.org/10.1016/j.preteyeres.2014.06.005

65. John W. “Integrative Omics Analysis of Rheumatoid Arthritis Identifies Non-Obvious Therapeutic Targets.” PLoS ONE 10.4 (2015): n. pag. Web.

66. Brugnera, E. "Unconventional Rac-Gef Activity Is Mediated through the Dock180-Elmo Complex." Nature Cell Biology 4.8 (2002): 574-82. Print.

67. Park, D. "Bai1 Is an Engulfment Receptor for Apoptotic Cells Upstream of the Elmo/Dock180/Rac Module." Nature 450.7168 (2007): 430-U10. Print.

68. R.Elliott, Michael. "Unexpected Requirement for Elmo1 in Apoptotic Germ Cell Clearance in Vivo." Nature 467.7313 (2010): 5. Print.

69. Stevenson, Catherine. "Essential Role of Elmo1 in Dock2-Dependent Lympocyte Migration." J.Immunol 198.12 (2014): 9. Print.

70. Chan, A. "The Gtpase Rac Regulates the Proliferation and Invasion of Fibroblast-Like Synoviocytes from Rheumatoid Arthritis Patients." *Molecular Medicine* 13.5-6 (2007): 297-304. Print.

71. Lefevre, S. "Synovial Fibroblasts Spread Rheumatoid Arthritis to Unaffected Joints." Nature Medicine 15.12 (2009): 1414-U10. Print.

72. Lee, D. M. "Cadherin-11 in Synovial Lining Formation and Pathology in Arthritis." Science 315.5814 (2007): 1006-10. Print.

73. Waterhouse, A. M., Procter, J. B., Martin, D. M. a, Clamp, M., & Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics, 25(9), 1189–1191. http://doi.org/10.1093/bioinformatics/btp033

74. Huang, Bingding. " LIGSITE^{csc} : Predicting Ligand Binding Sites Using the Connolly Surface and Degree of Conservation." *BMC Structual Biology* 6 19 (2006): 11. Print.

75. Fang, Yi. "Ben-Naim's "Pitfall": Don Quixote's Windmill." *Open Journal of Biophysics OJBIPHY* 3.1 (2013): 13-21. Print.

76. Johnson, David K. "Druggable Protein Interaction Sites Are More Predisposed to Surface Pocket Formation Than the Rest of the Protein Surface." *PLOS* 9 3 (2013): 10. Print.

77. Irwin, John J. "ZINC: A Free Tool to Discover Chemistry for Biology." Journal of Chemical Information and Modeling 2012: 1757–1768. Web.

78. Trott, Oleg. "AutoDock Vina Manual." AutoDock Vina. Molecular Graphics Lab at The Scripps Research Institute, 4 June 2009. Web. 7 Dec. 2015.

79. O'Boyle, Noel M. "Open Babel: An Open Chemical Toolbox." Journal of Cheminformatics 3.10 (2011): 1-14. Web.

80. Rogers, D. J., and T. T. Tanimoto. "A Computer Program for Classifying Plants." Science 132.3434 (1960): 1115-118. Print.

81. D.A. Case, J.T. Berryman, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman (2015), AMBER 2015, University of California, San Francisco.

82. Li, Nan, Richard I. Ainsworth, Bo Ding, Tingjun Hou, and Wei Wang. "Using Hierarchical Virtual Screening To Combat Drug Resistance of the HIV-1 Protease." *Journal of Chemical Information and Modeling J. Chem. Inf. Model.* (2015): 1400-412. Print.

83. Duan, Yong, Chun Wu, Shibasish Chowdhury, Mathew C. Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, James Caldwell, Junmei Wang, and Peter Kollman. "A Point-charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-phase Quantum Mechanical Calculations." J. Comput. Chem. Journal of Computational Chemistry (2003) 24.16: 1999-2012. Print.

84. Andersen, Hans C. “Molecular Dynamics Simulations at Constant Pressure And/or Temperature.” The Journal of Chemical Physics 72.1980 (1980): 2384. Web.

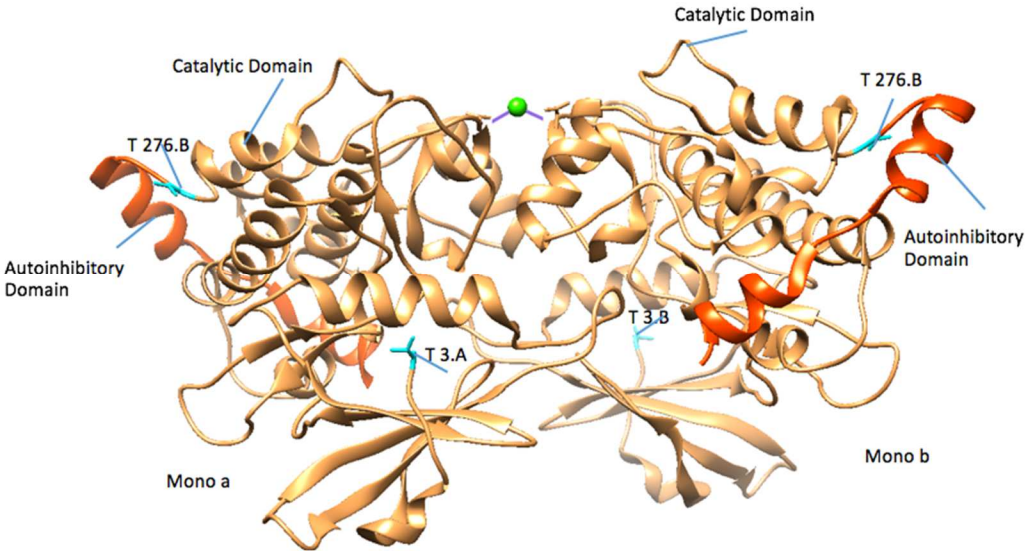
85. Case, David A. “The Amber Biomolecular Simulation Programs.” Journal of Computational Chemistry 2005: 1668–1688. Web.

86. Hou, Tingju. “Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations.” Journal of Chemical Information and Modeling 51.1 (2011): 69–82. Web.

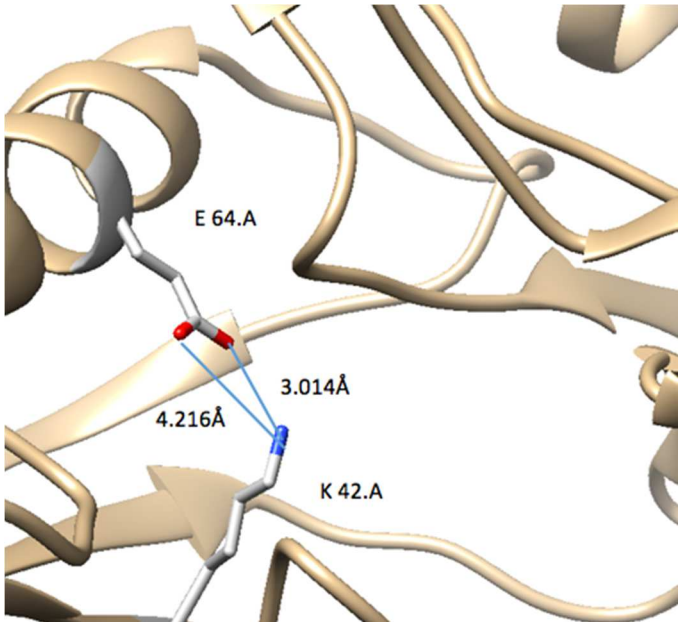
87. Ryckaert, J P, G Ciccotti, and H J C Berendsen. “Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes.” Journal of Computational Physics 23 (1977): 327–341. Web.

88. Darden, Tom, Darrin York, and Lee Pedersen. “Particle Mesh Ewald: An N□log(N) Method for Ewald Sums in Large Systems.” The Journal of Chemical Physics 98.12 (1993): 10089. Web.

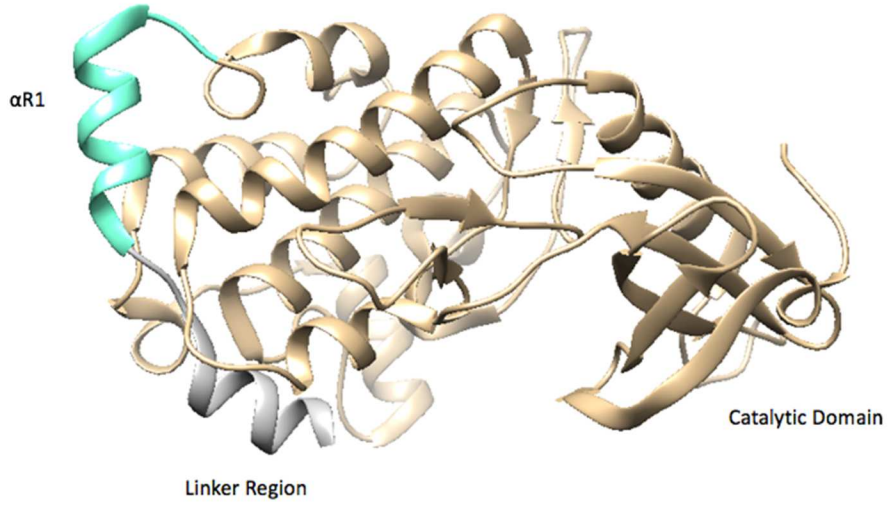
Supplementary Figures



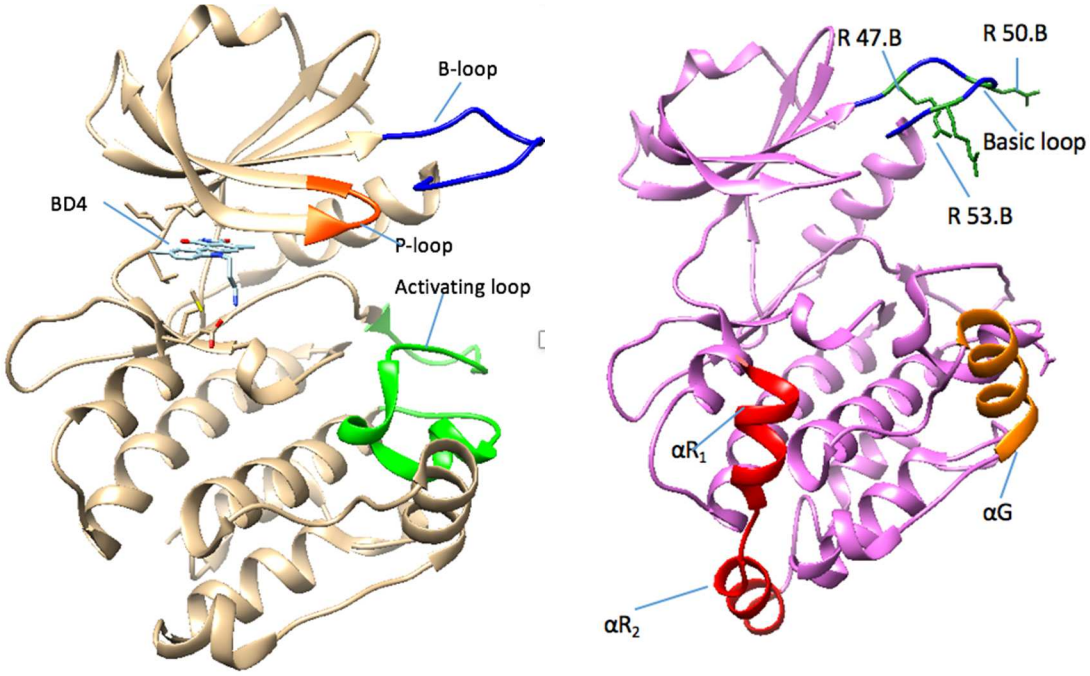
Sup. Figure A. shows the structure of the mDAPK2. The catalytic region is colored in sandy brown, a CAM binding domain (not shown, only the calcium ion is displayed in green), the autoinhibitory region is colored in orange red and the residues marking as the beginning and end of the catalytic domain (T.3 and T.276) are colored in cyan.



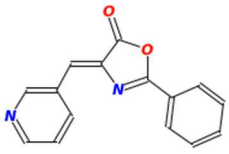
Sup. Figure B. shows the salt bridge between E64 and K42 (in light grey) at the active site of DAPK2, with the distance between side chain oxygen of E64 and nitrogen of K42 labeled



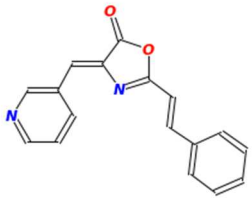
Sup. Figure C. shows the catalytic region (residue 3-274, in tan), autoinhibitory region of DAPK2 consisting of a linker region (residue 277-291, in aquamarine,) and the autoinhibitory helix αR1 (residue 292-301,in gray).



Sup. Figure D. 1. shows the structure of basic loop (in blue) and the basic residue Arg47, Arg50, and Arg53 (in green), between helices αG (in orange) and αR1 (in red).
Sup. Figure D. 2. displays the structure of the p-loop in DAPK1 (in blue), located between the basic loop (in orange red), and the activation loop (in green).

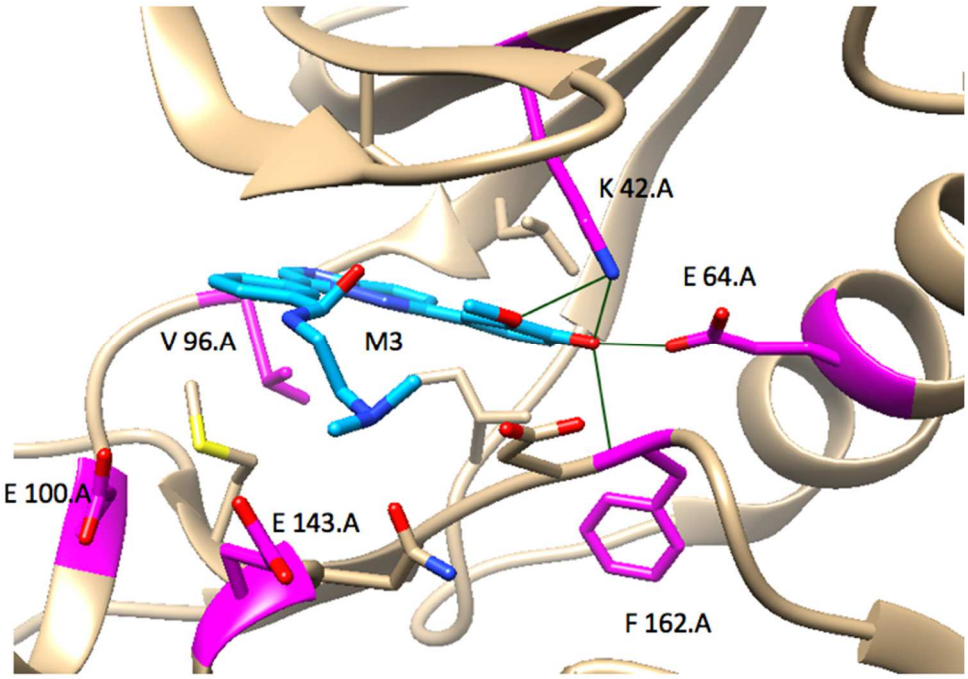


M1. 84% inhibition at 10μM for inhibition at 10μM DAPK1 and 71% inhibition at 10μM for DAPK3

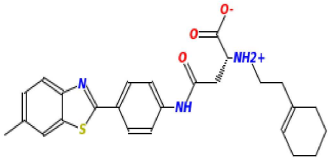


M2. 91% inhibition at 10μM for for DAPK1 and 100% inhibition for for DAPK3

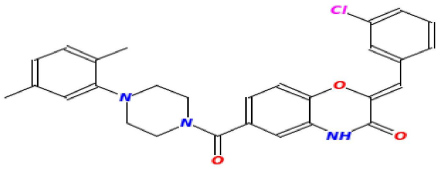
Sup. Figure E. shows the structures of M1 and M2, which are the most potent inhibitors for DAPK1 and 3 respectively.



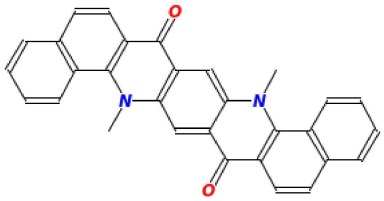
Sup. Figure F. shows the structure of the active site for M3 (cyan) binding to DAPK1 at the active site, along with the pocket residues (magenta) around the ligand.



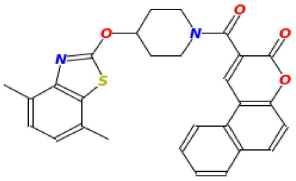
1.ZINC72438491 (-12.0Kcal/mol)



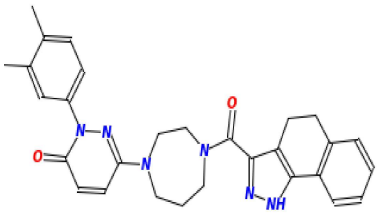
2.ZINC32983948 (-11.9Kcal/mol)



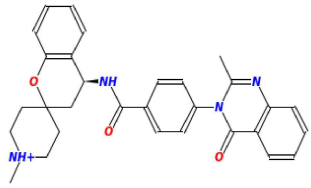
3. ZINC22007970 (-11.9Kcal/mol)



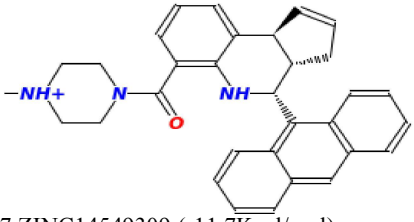
4.ZINC49442684 (-11.8Kcal/mol)



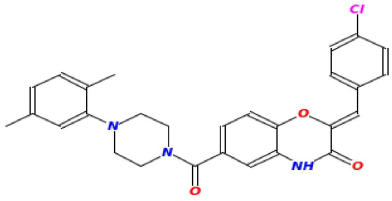
5.ZINC15730186(-11.8Kcal/mol)



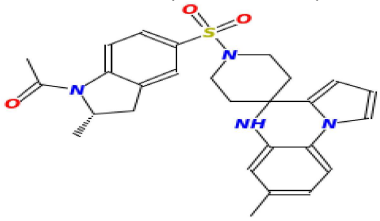
6.ZINC32721460(-11.7Kcal/mol)



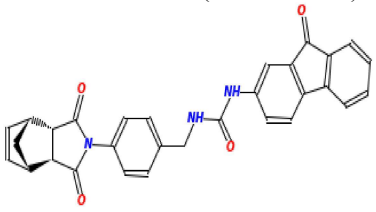
7.ZINC14549309 (-11.7Kcal/mol)



8.ZINC32983502 (-11.7Kcal/mol)

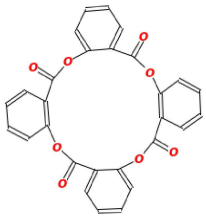


9.ZINC21152557 (-11.7Kcal/mol)

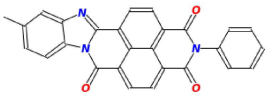


10.ZINC72438494 (-11.7Kcal/mol)

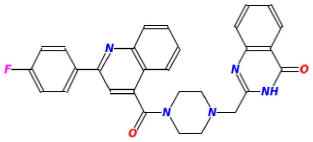
Sup. Figure G. Top 10 ligand-candidates for Rigid Docking at DAPK2 (exhaustiveness=1). All of the ligands were ranked based on their binding affinity, numbered from 1 to 10.



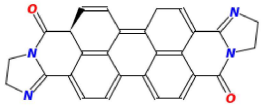
1.ZINC05015110
(-13.3Kcal/mol,480.428g/mol)



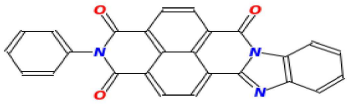
2.ZINC00626857
(-12.5Kcal/mol, 429.435g/mol)



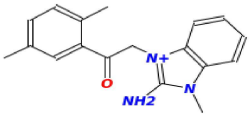
3.ZINC24527931
(-12.5Kcal/mol,493.542g/mol)



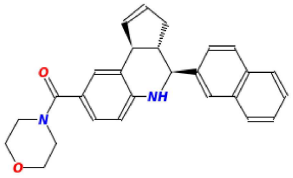
4. ZINC22013692
(-12.4Kcal/mol,442.478g/mol)



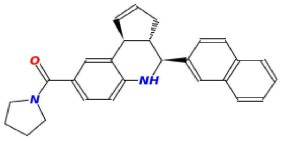
5.ZINC00988780
(-12.3Kcal/mol,415.408g/mol)



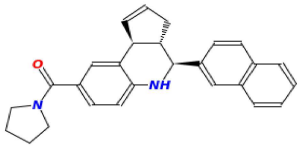
6.ZINC05330647
(-12.3Kcal/mol,367.408g/mol)



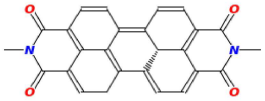
7.ZINC14475557
(-12.3Kcal/mol,394.518g/mol)



8.ZINC14475559
(-12.3Kcal/mol, 394.518g/mol)



9.ZINC22007970
(-12.3Kcal/mol, 440.502 g/mol)



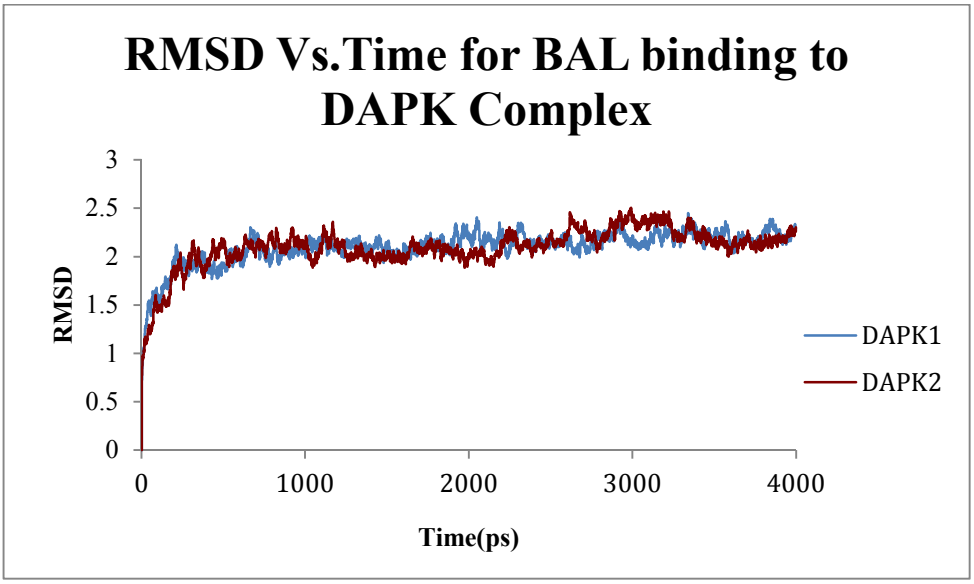
10.ZINC71257152
(-12.3Kcal/mol, 420.424 g/mol)

Sup. Figure H. shows the top 10 ligand-candidates for DAPK1 in flexible docking. These ligands were numbered and ranked based on the binding affinity only, with exhaustiveness=9. As indicated by Figure D., all of the ligand candidates produced by flexible docking is larger than 350g/mol. Based on their binding affinity difference produced by DAPK1 and 2, ZINC00988780 was selected to perform MD simulation described in the next section.

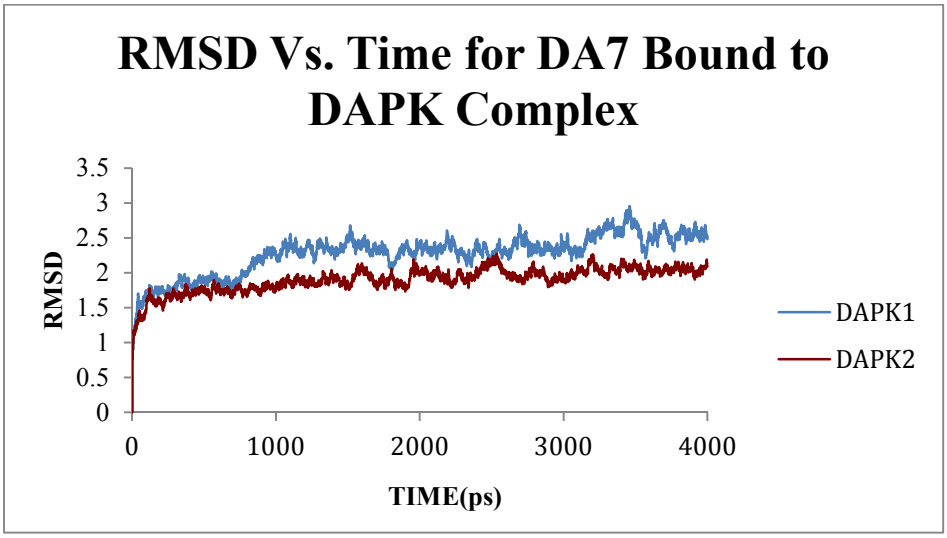
Sup. Table 1. shows ligand-based screening results for the top 1,000-ligand candidates for DAPK2. The *Cluster#* column stands for highest rank of the ligand in each cluster and *ZINC ID* column stands for the ZINC ID of the ligands in each numbered cluster. In the *ZINC ID* column, the ZINC ID colored in orange is the ligand with highest binding affinity which is also the one that serve as the reference molecule in their corresponding cluster. In this table, four clusters with the largest number of candidates were included, which were selected from 731 unique clusters constructed from R. Based on the screening results, it can be indicated that the structures of the top 1,000 ligands were not highly similar to each other. However, since the molecular weight of these top 1,000 ligand candidates were all larger than 380g/mol, the ligand candidates were filtered again, according to their molecular weight.

Cluster #	ZINC ID
64	ZINC09293407, ZINC35103583, ZINC09282744, ZINC09282657, ZINC09282661, ZINC09309355, ZINC33516805, ZINC09309697, ZINC04090333, ZINC33725605, ZINC09235851, ZINC09309613, ZINC09293469, ZINC02425080, ZINC09232956, ZINC33600171, ZINC09292086
122	ZINC72472326, ZINC72472332, ZINC72472312, ZINC72472331, ZINC72472314, ZINC72472310, ZINC72472316, ZINC72472322, ZINC72472313, ZINC72472311, ZINC72472315, ZINC72472318, ZINC72472321
29	ZINC72461073, ZINC72461075, ZINC72461079, ZINC72461074, ZINC72461071, ZINC72461072, ZINC72461077, ZINC72461083, ZINC72461076, ZINC72461078, ZINC72461070
78	ZINC72472410, ZINC72472586, ZINC72472411, ZINC72472412, ZINC72472587

(1)



(2)

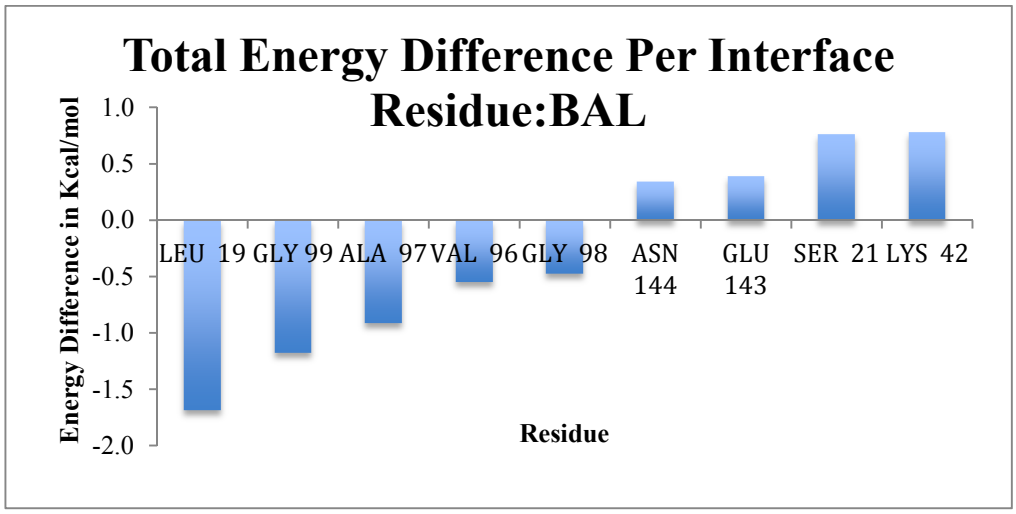


Sup. Figure I. 1-2. shows the RMSD vs. Time plot for each selected ligand bound to DAPK1 and 2 respectively.

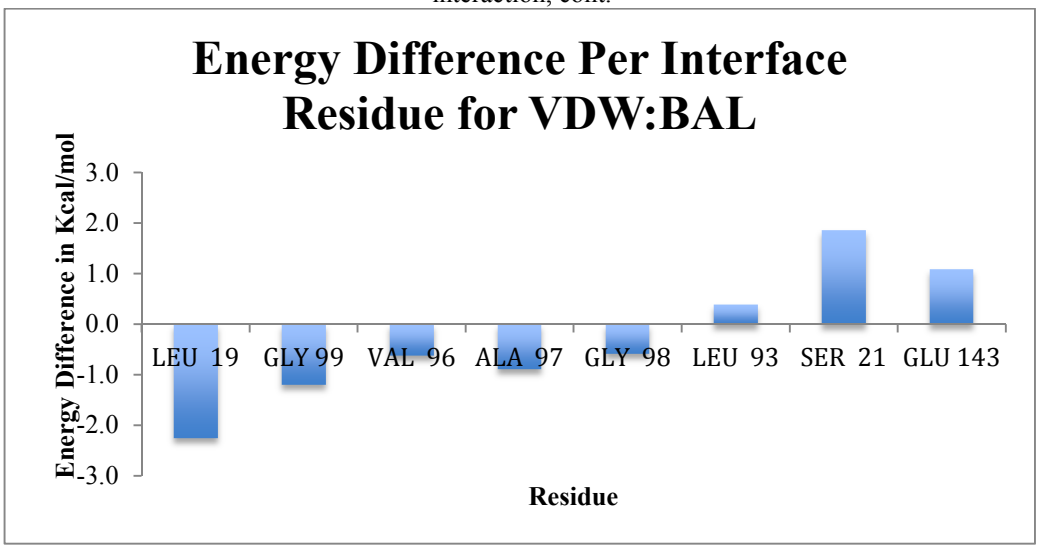
Sup. Figure I. 1. RMSD vs. Time for BAL bound to DAPK1 and 2 respectively.

Sup. Figure I. 2. RMSD vs. Time for DA7 bound to DAPK1 and 2 respectively.

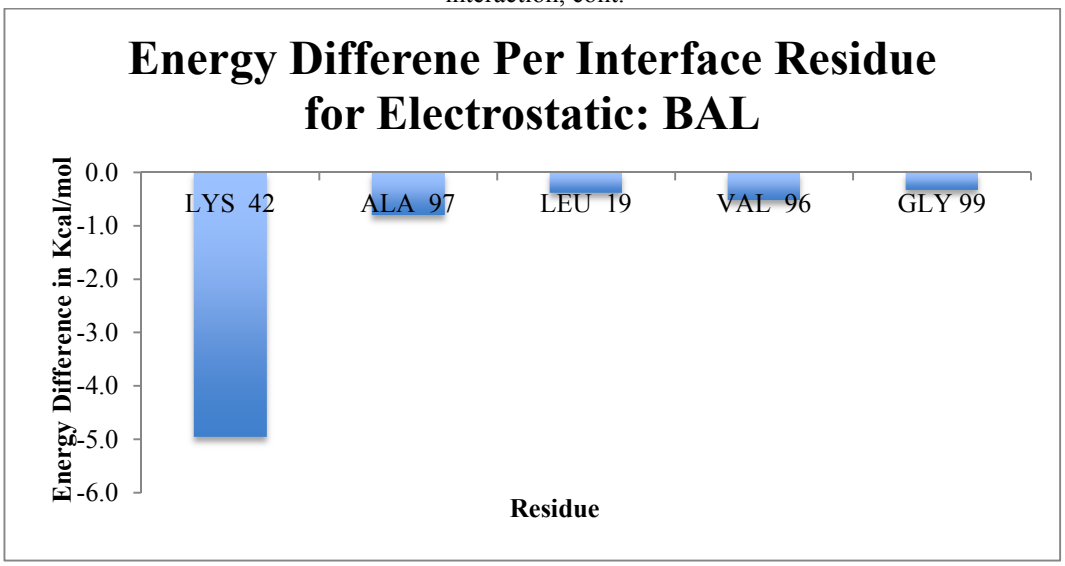
Sup. Table 2. a. shows the energy difference per residue between DAPK1 and 2 for BAL, of total energy, cont.



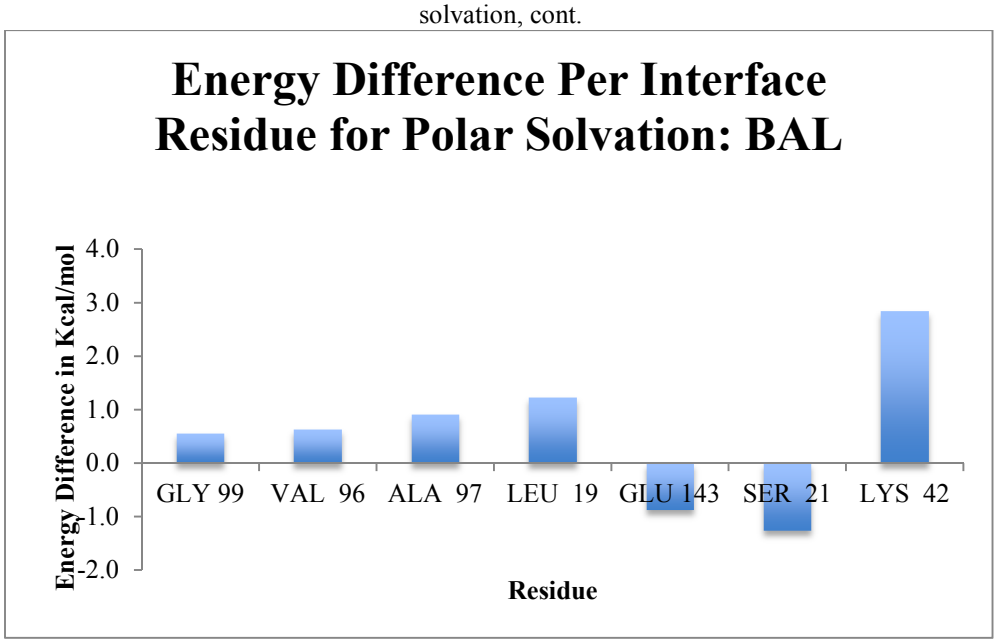
Sup. Table 2. b. shows the energy difference per residue between DAPK1 and 2 for BAL, of VDW interaction, cont.



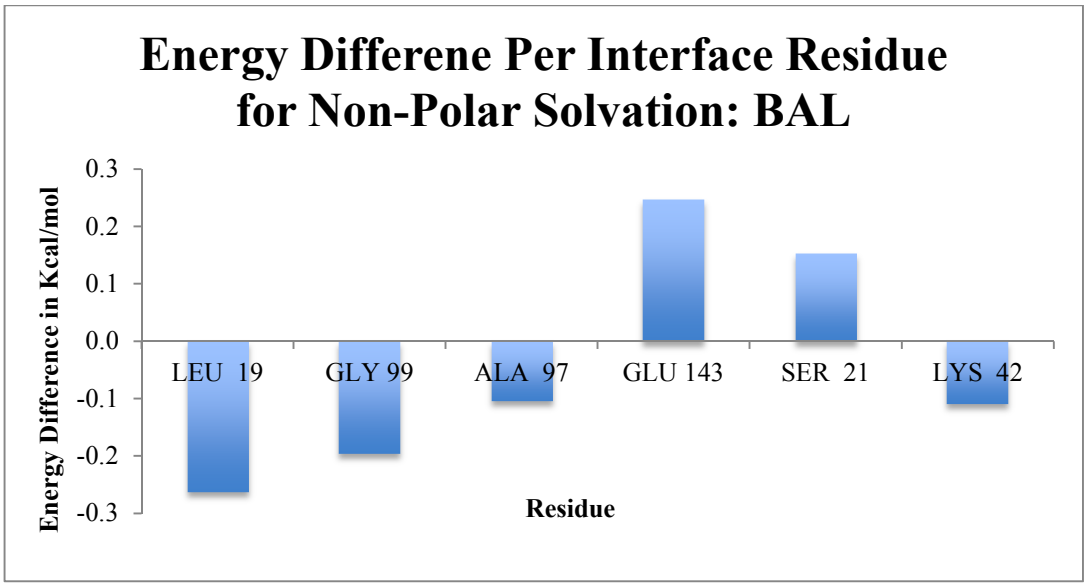
Sup. Table 2. c. shows the energy difference per residue between DAPK1 and 2 for BAL, of electrostatic interaction, cont.



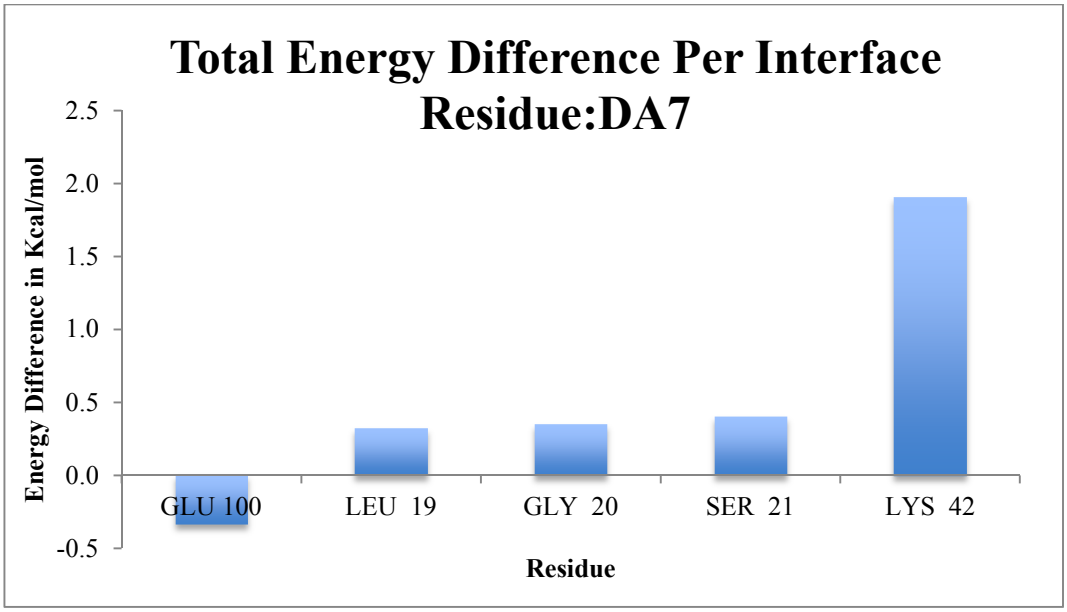
Sup. Table 2. d. shows the energy difference per residue between DAPK1 and 2 for BAL, of polar solvation, cont.



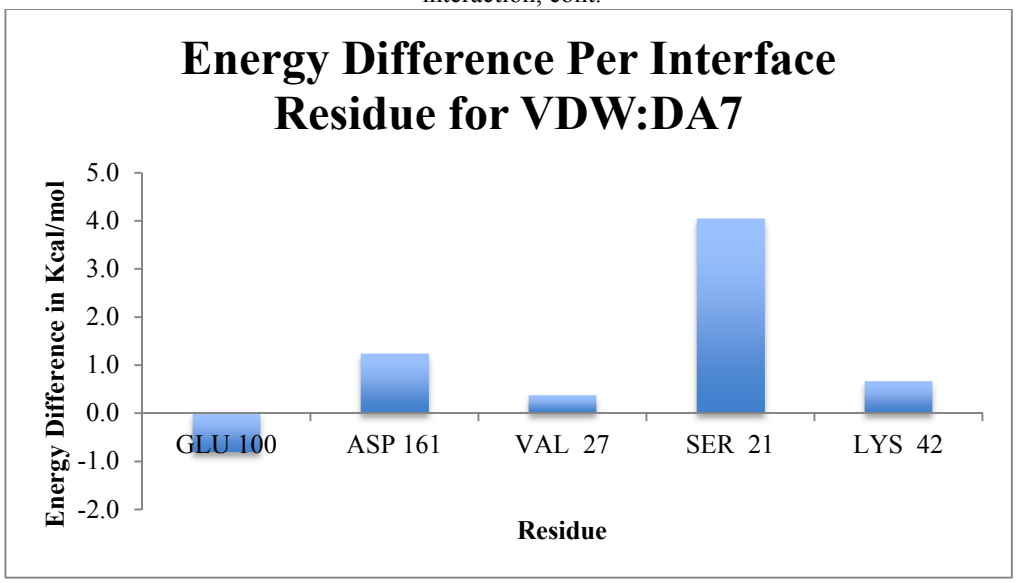
Sup. Table 2. e. shows the energy difference per residue between DAPK1 and 2 for BAL, of non-polar solvation.



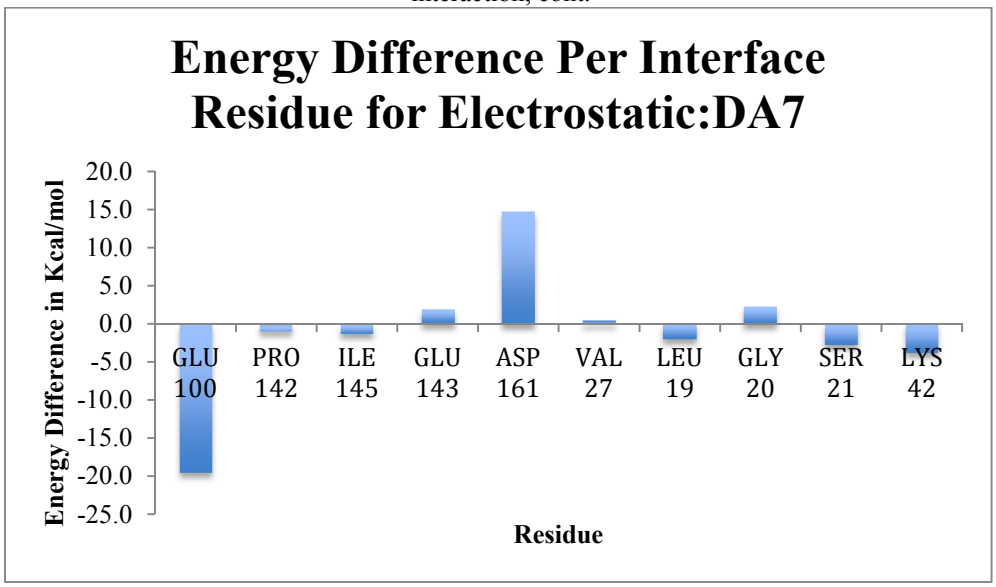
Sup. Table 3. a. shows the energy difference per residue between DAPK1 and 2 for DA7, of total energy, cont.



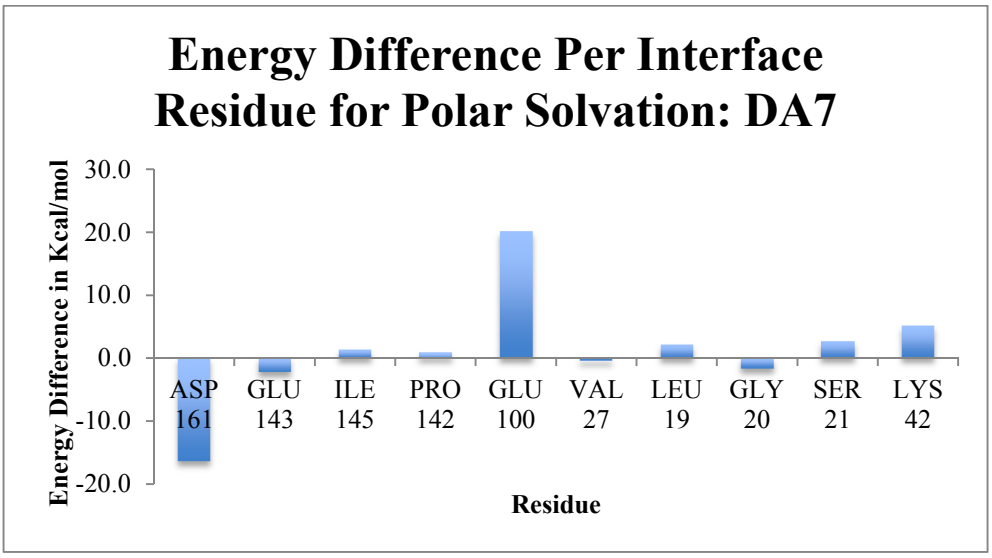
Sup. Table 3. b. shows the energy difference per residue between DAPK1 and 2 for DA7, of VDW interaction, cont.



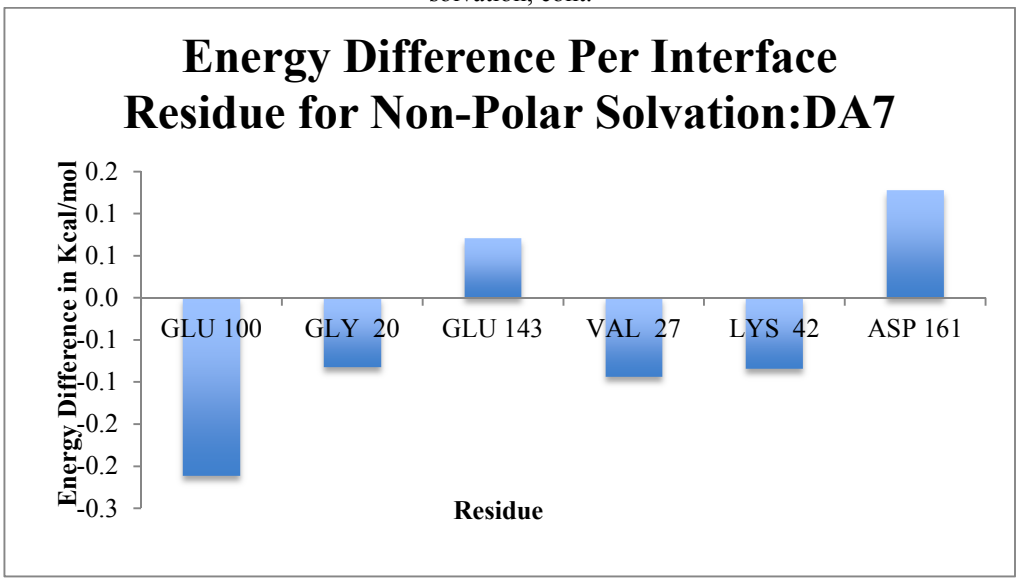
Sup. Table 3. c. shows the energy difference per residue between DAPK1 and 2 for DA7, of electrostatic interaction, cont.



Sup. Table 3. d. shows the energy difference per residue between DAPK1 and 2 for DA7, of polar solvation, cont.



Sup. Table 3. e. shows the energy difference per residue between DAPK1 and 2 for DA7, of non-polar solvation, cont.



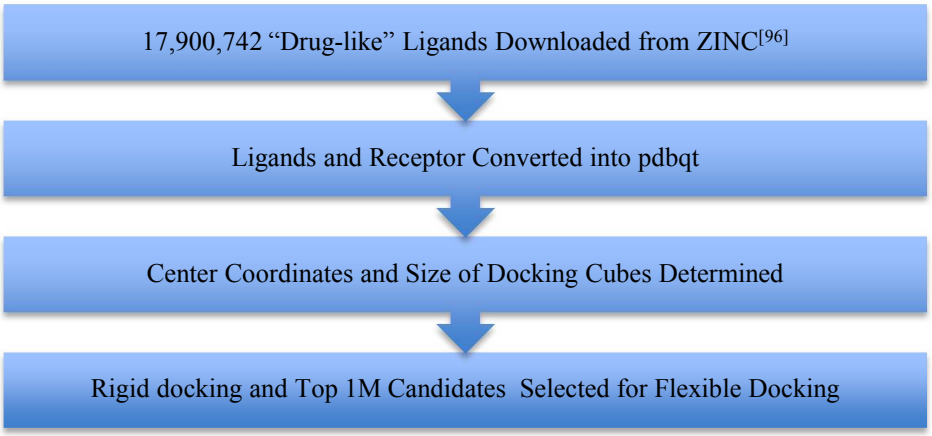
Sup. Table 4. shows the complete form shows the complete form for the total energy difference between DAPK1 and 2 when bound to BAL and DA7 respectively.

Total Energy Difference for BAL Binding to DAPK1 and 2				
	DAPK1-BAL	DAPK2-BAL	Difference	Absolute Value for Difference
LEU 18	-2.0726	-0.3933	-1.6793	1.6793
GLY 98	-1.1653	0.0056	-1.1709	1.1709
ALA 96	-0.8834	0.0229	-0.9063	0.9063
VAL 95	-0.8519	-0.3083	-0.5437	0.5437
GLY 97	-0.4585	0.0067	-0.4652	0.4652
GLY 19	-1.111011464	-0.897753302	-0.213258162	0.2133
GLU 99	0.122879062	0.316784936	-0.193905874	0.1939
GLU 93	-0.138014406	-0.069926406	-0.068088	0.0681
ILE 144	0.013542	-0.004934	0.018476	0.0185
LEU 94	-0.065712	-0.091845	0.026133	0.0261
PRO 141	0.000906	-0.026088	0.026994	0.0270
VAL 26	-1.692250136	-1.719859914	0.027609778	0.0276
LYS 28	0.06106772	0.01931	0.04175772	0.0418
ASP 160	-0.35676521	-0.560772952	0.204007742	0.2040
LEU 92	-0.572189018	-0.869804144	0.297615126	0.2976
ASN 143	-0.0124	-0.3489	0.3365	0.3365
GLU 142	0.1701	-0.2208	0.3908	0.3908
SER 20	-0.1910	-0.9509	0.7599	0.7599
LYS 41	0.2000	-0.5813	0.7813	0.7813
			-2.3296	
Total Energy Difference for DA7 Binding to DAPK1 and 2				
	DAPK1-DA7	DAPK2-DA7	Diff-DA7	diff-abs
GLU 100	-0.3308	0.0012	-0.3319	0.3319
ASP 161	0.01381	0.267704464	-0.253894464	0.2539
GLU 143	-0.018716541	0.148710258	-0.167426798	0.1674
PRO 142	-0.001848	0.007682	-0.00953	0.0095
ILE 145	0.015412882	0.01381	0.001602882	0.0016
GLY 98	-0.00422	-0.022383901	0.018163901	0.0182
ALA 97	0.003402283	-0.049596501	0.052998784	0.0530
ASN 144	-0.119563182	-0.182110422	0.06254724	0.0625
LEU 95	-0.009121915	-0.075923632	0.066801717	0.0668
GLU 94	-0.069397173	-0.152151659	0.082754486	0.0828
LYS 29	0.167652	0.083176	0.084476	0.0845
GLY 99	0.138270298	-0.01769776	0.155968058	0.1560
VAL 96	-0.311754403	-0.488160182	0.176405779	0.1764
LEU 93	-0.3197176	-0.505672374	0.185954774	0.1860
VAL 27	-1.678768482	-1.962495618	0.283727136	0.2837
LEU 19	-0.6773	-1.0001	0.3229	0.3229
GLY 20	-0.6756	-1.0289	0.3532	0.3532
SER 21	-0.1271	-0.5251	0.3980	0.3980
LYS 42	0.6842	-1.2208	1.9050	1.9050
			3.3877	4.9132

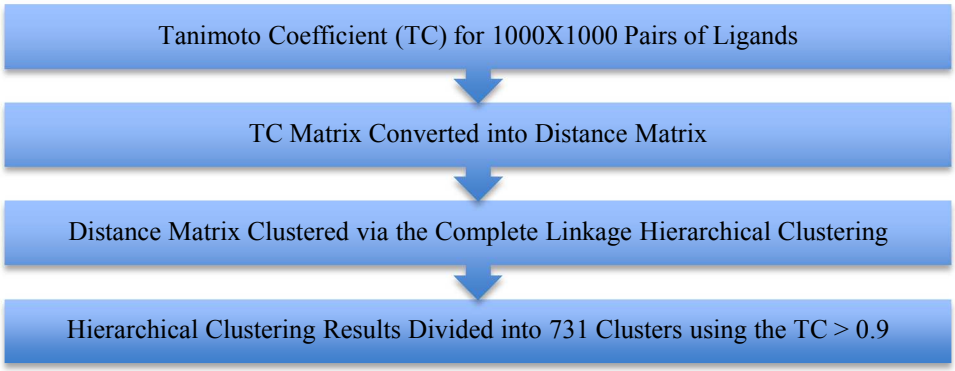
Sup. Table 5. is the complete table for energy decomposition data.

DAPK1-BAL																
	Location	van der Waals			Electrostatic			Polar Solvation			Non-Polar Solv.			TOTAL		
LYS 41	R LYS 41	-1.41002	0.3570134	0.015966124	-3.171106	1.64251	0.07346	4.99072	2.19347	0.0981	-0.209625	0.03599	0.00161	0.1999667	0.9648	0.04315
GLU 142	R GLU 142	-0.135062	0.0325238	0.001454509	-0.415798	0.22094	0.00988	0.72102	0.23836	0.01066	-0.000101	0.0008	3.56E-05	0.1700606	0.03936	0.00176
GLU 99	R GLU 99	-0.567756	0.1748755	0.007820668	0.09157	0.53566	0.02396	0.65476	0.53929	0.02412	-0.055695	0.04012	0.00179	0.1228791	0.23396	0.01046
LYS 28	R LYS 28	-0.062854	0.0199278	0.000891198	0.003916	0.11078	0.00495	0.12009	0.12922	0.00578	-8.03E-05	0.00118	5.27E-05	0.0610677	0.0474	0.00212
ILE 144	R ILE 144	-0.062856	0.009858	0.000440861	-0.018522	0.02564	0.00115	0.09492	0.04356	0.00195	0	0	0	0.013542	0.02389	0.00107
PRO 141	R PRO 141	-0.01766	0.0030972	0.000138509	-0.015142	0.01621	0.00072	0.03371	0.0179	0.0008	0	0	0	0.000906	0.00554	0.00025
ASN 143	R ASN 143	-0.111492	0.0185092	0.000827756	0.044274	0.03374	0.00151	0.05482	0.0421	0.00188	0	0	0	-0.012402	0.02959	0.00132
LEU 94	R LEU 94	-0.12362	0.0277523	0.001241119	0.03973	0.03297	0.00147	0.01818	0.03264	0.00146	0	0	0	-0.065712	0.03054	0.00137
GLU 93	R GLU 93	-0.09826	0.019929	0.000891251	0.48957	0.21548	0.00964	-0.52928	0.19115	0.00855	-4.04E-05	0.0003	1.35E-05	-0.1380144	0.03983	0.00178
SER 20	R SER 20	-0.405678	0.2018404	0.009026576	-0.227624	0.33634	0.01504	0.51929	0.39702	0.01776	-0.077028	0.04831	0.00216	-0.1910398	0.13445	0.00601
ASP 160	R ASP 160	-0.76127	0.220137	0.009844807	-0.18441	0.4653	0.0208	0.6774	0.5173	0.0231	-0.08849	0.0282	0.0013	-0.35677	0.2538	0.0113
GLY 97	R GLY 97	-0.59189	0.179885	0.008044684	-0.04301	0.0743	0.0033	0.1892	0.0731	0.0033	-0.01283	0.0088	0.0004	-0.45853	0.1489	0.0067
LEU 92	R LEU 92	-0.51522	0.301691	0.013492015	-0.09231	0.0469	0.0021	0.109	0.0379	0.0017	-0.07369	0.0207	0.0009	-0.57219	0.3074	0.0137
VAL 95	R VAL 95	-1.09274	0.289405	0.012942591	-0.46814	0.251	0.0112	0.8039	0.1538	0.0069	-0.09491	0.0181	0.0008	-0.85194	0.2979	0.0133
ALA 96	R ALA 96	-0.9187	0.227542	0.010175979	-0.79665	0.3748	0.0168	0.9363	0.244	0.0109	-0.10438	0.0233	0.001	-0.88342	0.3377	0.0151
GLY 19	R GLY 19	-1.17034	0.251368	0.01124154	0.04135	0.1546	0.0069	0.1814	0.1293	0.0058	-0.1634	0.0226	0.001	-1.11101	0.2919	0.0131
GLY 98	R GLY 98	-1.21795	0.267651	0.01196972	-0.36192	0.2786	0.0125	0.6098	0.1692	0.0076	-0.19524	0.0274	0.0012	-1.16534	0.256	0.0114
VAL 26	R VAL 26	-1.71368	0.251725	0.011257467	0.06893	0.0566	0.0025	0.2	0.0696	0.0031	-0.24745	0.0311	0.0014	-1.69225	0.2521	0.0113
LEU 18	R LEU 18	-2.76859	0.44805	0.020037413	-0.51097	0.5766	0.0258	1.5223	0.4308	0.0193	-0.31536	0.0471	0.0021	-2.0726	0.5064	0.0226
DAPK1-DA7																
LYS 42	R LYS 41	-1.215284	0.4446608	0.019885837	16.271336	1.17104	0.05237	-14.1025	1.19921	0.05363	-0.269374	0.04249	0.0019	0.6841803	0.54124	0.0242
LYS 29	R LYS 28	-0.029128	0.0068548	0.000306554	11.937522	1.02978	0.04605	-11.7407	1.004	0.0449	0	0	0	0.167652	0.0225	0.00101
GLY 99	R GLY 98	-0.298148	0.1574228	0.007040162	2.025888	0.54368	0.02431	-1.53446	0.41894	0.01874	-0.055014	0.02927	0.00131	0.1382703	0.19082	0.00853
ASP 161	R ASP 160	-1.062268	0.2012719	0.009001155	-14.25622	0.97157	0.04345	15.5736	1.13706	0.05085	-0.235638	0.03365	0.0015	0.0195219	0.39146	0.01751
ILE 145	R ILE 144	-0.09388	0.0321636	0.0014384	-1.40731	0.4972	0.02224	1.51791	0.54401	0.02433	-0.001309	0.00301	0.00013	0.0154129	0.02642	0.00118
ALA 97	R ALA 96	-0.050802	0.0237616	0.001062652	-0.114974	0.33524	0.01499	0.16918	0.33121	0.01481	-5.72E-06	9.52E-05	4.26E-06	0.0034023	0.00891	0.0004
PRO 142	R PRO 141	-0.029448	0.0074958	0.000335223	-0.927542	0.1852	0.00828	0.95514	0.19276	0.00862	0	0	0	-0.001848	0.0064	0.00029
LEU 95	R LEU 94	-0.051184	0.0172176	0.000769995	0.506096	0.07104	0.00318	-0.46403	0.06448	0.00288	-1.92E-06	4.28E-05	1.91E-06	-0.0091219	0.01213	0.00054
GLU 143	R GLU 142	-0.32981	0.0977446	0.004371271	-20.46315	2.23546	0.09997	20.8341	2.27262	0.10163	-0.059825	0.03352	0.0015	-0.0187165	0.09112	0.00407
GLU 94	R GLU 93	-0.026978	0.0080728	0.000361025	-9.800872	0.36536	0.01634	9.7585	0.37101	0.01659	-4.52E-05	0.00033	1.47E-05	-0.0693972	0.00925	0.00041
ASN 144	R ASN 143	-0.152788	0.0373415	0.001669965	-0.020156	0.22113	0.00989	0.05367	0.22722	0.01016	-0.000291	0.00175	7.81E-05	-0.1195632	0.05069	0.00227
GLY 98	R GLY 97	-0.03701	0.0144558	0.000646483	0.352698	0.20432	0.00914	-0.31991	0.18967	0.00848	0	0	0	-0.00422	0.01255	0.00056
SER 21	R SER 20	-0.57938	0.2736637	0.012238613	-2.42475	0.58197	0.02603	3.00761	0.75476	0.03375	-0.13057	0.06269	0.0028	-0.1270884	0.13193	0.0059
VAL 96	R VAL 95	-0.366462	0.1425736	0.006376087	-0.12174	0.15797	0.00706	0.22273	0.16765	0.0075	-0.046278	0.01994	0.00089	-0.3117544	0.13043	0.00583
LEU 93	R LEU 92	-0.277234	0.1009761	0.004515787	0.116978	0.09266	0.00414	-0.10771	0.09267	0.00414	-0.051754	0.02273	0.00102	-0.3197176	0.1138	0.00509
GLU 100	R GLU 99	-0.880884	0.280313	0.012535979	-28.47859	3.73928	0.16723	29.2389	4.16795	0.1864	-0.210245	0.03342	0.00149	-0.3307788	0.58528	0.02617
GLY 20	R GLY 19	-0.938608	0.2999906	0.013415986	1.341222	0.35276	0.01578	-0.9088	0.2624	0.01174	-0.169451	0.01991	0.00089	-0.6756335	0.34693	0.01552
LEU 19	R LEU 18	-1.076774	0.3100532	0.013866002	-2.215108	0.5166	0.0231	2.7497	0.65451	0.02927	-0.135068	0.04037	0.00181	-0.6772504	0.24852	0.01111
VAL 27	R VAL 26	-1.381708	0.3354865	0.015003411	0.171356	0.14557	0.00651	-0.2372	0.13976	0.00625	-0.231218	0.03534	0.00158	-1.6787685	0.37044	0.01657
DAPK2-BAL																
VAL 25	R VAL 25	-1.58405	0.430103	0.019234812	-0.04681	0.0441	0.002	0.1466	0.076	0.0034	-0.23557	0.0402	0.0018	-1.71986	0.425	0.019
SER 19	R SER 19	-2.25377	0.396254	0.017720997	-0.23634	0.6483	0.029	1.7683	0.4891	0.0219	-0.22913	0.0353	0.0016	-0.95094	0.5256	0.0235
GLY 18	R GLY 18	-1.40931	0.306693	0.013715718	-0.02035	0.3436	0.0154	0.7449	0.4422	0.0198	-0.213	0.0363	0.0016	-0.89775	0.3596	0.0161
LEU 91	R LEU 91	-0.90047	0.244862	0.010950561	0.03924	0.0709	0.0032	0.0586	0.0599	0.0027	-0.06719	0.0185	0.0008	-0.8698	0.2642	0.0118
LYS 40	R LYS 40	-1.39347	0.291829	0.013051009	-1.23137	0.9787	0.0438	2.1441	0.9647	0.0431	-0.10059	0.0226	0.001	-0.58131	0.4731	0.0212
ASP 159	R ASP 159	-1.47152	0.378083	0.016908394	-1.21918	0.5452	0.0244	2.3418	0.6797	0.0304	-0.21188	0.0282	0.0013	-0.56077	0.5539	0.0248
LEU 17	R LEU 17	-0.51817	0.20917	0.009354361	-0.13296	0.1762	0.0079	0.3102	0.143	0.0064	-0.05237	0.0357	0.0016	-0.39327	0.2219	0.0099
ASN 142	R ASN 142	-0.40635	0.140767	0.006295297	0.00609	0.0726	0.0032	0.0626	0.0865	0.0039	-0.01127	0.0123	0.0006	-0.34895	0.181	0.0081
VAL 94	R VAL 94	-0.486406	0.1235631	0.005525909	0.05418	0.12001	0.00537	0.1744	0.11702	0.00523	-0.050427	0.01305	0.00058	-0.3082514	0.12053	0.00539
GLU 141	R GLU 141	-1.213846	0.6274218	0.028059157	-0.35239	0.44871	0.02007	1.59153	0.95516	0.04272	-0.246055	0.12478	0.00558	-0.2207625	0.45596	0.02039
LEU 93	R LEU 93	-0.16702	0.0610606	0.002730714	-0.0392	0.05529	0.00247	0.11536	0.06028	0.0027	-0.000981	0.00342	0.00015	-0.091845	0.05801	0.00259
GLU 92	R GLU 92	-0.290468	0.1296488	0.00579807	-0.84531	0.18377	0.00822	1.07543	0.24979	0.01117	-0.00958	0.00617	0.00028	-0.0699264	0.19991	0.00894
PRO 140	R PRO 140	-0.04903	0.0180477	0.00080712	-0.040518	0.01981	0.00089	0.06346	0.0263	0.00118	0	0	0	-0.026088	0.01363	0.00061
ILE 143	R ILE 143	-0.05436	0.0129979	0.000581285	-0.014964	0.02461	0.0011	0.06439	0.04665	0.00209	0	0	0	-0.004934	0.01916	0.00086
GLY 97	R GLY 97	-0.0281	0.0068826	0.000307799	-0.034068	0.01839	0.00082	0.06772	0.0217	0.00097	0	0	0	0.00555	0.00973	0.00043
GLY 96	R GLY 96	-0.009248	0.0021453	9.59E-05	0.008272	0.01057	0.00047	0.00767	0.01124	0.0005	0	0	0	0.006698	0.00396	0.00018
LYS 27	R LYS 27	-0.041126	0.0092981	0.000415822	-0.242434	0.09089	0.00406	0.30287	0.09722	0.00435	0	0	0	0.01931	0.00918	0.00041
SER 95	R SER 95	-0.025686	0.0093263	0.000417084	0.000552	0.02713	0.00121	0.04805	0.03263	0.00146	0	0	0	0.022912	0.01109	0.0005
GLU 98	R GLU 98	-0.163144	0.0560845	0.002508174	-0.341468	0.3221	0.0144	0.84793	0.40895	0.01829	-0.026531	0.02824	0.00126	0.3167849	0.0891	0.00398
DAPK2-DA7																
ASP 159	R ASP 159	-2.288556	0.5483662	0.024523684	-29.01614	1.52211	0.06807	31.9356	1.25437	0.0561	-0.363164	0.0258	0.00115	0.2677045	0.7332	0.03279
GLU 141	R GLU 141	-0.47606	0.1102412	0.004930136	-22.31704	2.10973	0.09435	23.0722	2.20067	0.09842	-0.130412	0.03581	0.0016	0.1487103	0.1007	0.0045
LYS 27	R LYS 27	-0.029176	0.0051989	0.000232504	6.349464	0.15874	0.0071	-6.23711	0.15204	0.0068	0	0	0	0.083176	0.00634	0.00028
ILE 143	R ILE 143	-0.06856	0.0123873	0.000553979	-0.099974	0.09341	0.00418	0.18234	0.10764	0.00481	0	0	0	0.01381	0.01536	0.00069
PRO 140	R PRO 140	-0.023056	0.002604	0.000116455	-0.015424	0.04639	0.00207	0.04616	0.04509	0.00202	0	0	0	0.007682	0.00474	0.00021
GLU 98	R GLU 98	-0.069774	0.0209783	0.00093818	-9.073228	0.45346	0.02028	9.14424	0.46374	0.02074	-8.67E-05	0.00086	3.85E-05	0.0011513	0.02604	0.00116
GLY 97	R GLY 97	-0.053092	0.0179414	0.000802366	0.061414	0.06796	0.00304	-0.02281	0.07469	0.00334	-0.00321	0.00712	0.00032	-0.0176978	0.0236	0.00106
GLY 96	R GLY 96	-0.029006	0.0080631	0.000360594	0.163896	0.04092	0.00183	-0.15727	0.04144	0.00185	-1.90E-06	4.25E-05	1.90E-06	-0.0223839	0.01272	0.00057
SER 95	R SER 95	-0.156648	0.0682836	0.003053737	0.189896	0.17132	0.00766	-0.07415	0.16125	0.00721	-0.008691	0.00976	0.00044	-0.0495965	0.05738	0.00257
LEU 93	R LEU 93	-0.093088	0.0283766	0.001269041	0.143398	0.06045	0.0027	-0.12623	0.04775	0.00214	-7.63E-06	0.00013	6.03E-06	-0.0759236	0.02274	0.00102
GLU 92	R GLU 92	-0.098968	0.0591863	0.002646894	-8.532436	0.24574	0.01099	8.48212	0.2507	0.01121	-0.002868	0.00411	0.00018	-0.1521517	0.0586	0.00262
ASN 142	R ASN 142	-0.260432	0.0298487	0.001334875	0.169732	0.22446	0.01004	-0.0646	0.22296	0.00997	-0.026808	0.0122	0.00055	-0.1821104	0.03488	0.00156
VAL 94	R VAL 94	-0.584146	0.250824	0.011217191	-0.34747	0.19585	0.00876	0.50044	0.16045	0.00718	-0.056982	0.02016	0.0009	-0.4881602	0.24914	0.01114
LEU 91	R LEU 91	-0.466328	0.2227467	0.009961534	0.225038	0.04105	0.00184	-0.20495	0.03752	0.00168	-0.059436	0.02531	0.00113	-0.5056724	0.23018	0.01029
SER 19	R SER 19	-1.024346	0.3978166	0.017790901	0.298204	0.49442	0.02211	0.28462	0.25449	0.01138	-0.083533	0.04113	0.00184	-0.5250511	0.35724	0.01598
LEU 17	R LEU 17	-1.243088	0.2905726	0.0129948	-0.190902	0.17055	0.00763	0.57188	0.30038	0.01343	-0.138021	0.04196	0.00188	-1.0001329	0.29541	0.01321
GLY 18	R GLY 18	-0.815354	0.2291093	0.010246079	-0.8852	0.31325	0.01401	0.75918	0.21271	0.00951	-0.087487	0.02241	0.001	-1.0288634	0.31965	0.0143
LYS 40	R LYS 40	-1.890554	0.3413612	0.015266137	20.0607	0.86708	0.03878	-19.2047	0.70191	0.03139	-0.186279	0.03038	0.00136	-1.2207914	0.47161	0.02109
VAL 25	R VAL 25	-1.75063	0.3090717	0.013822104	-0.238064	0.09707	0.00434	0.2545	0.07033	0.00315	-0.228302	0.02453	0.0011	-1.9624956	0.30445	0.01362

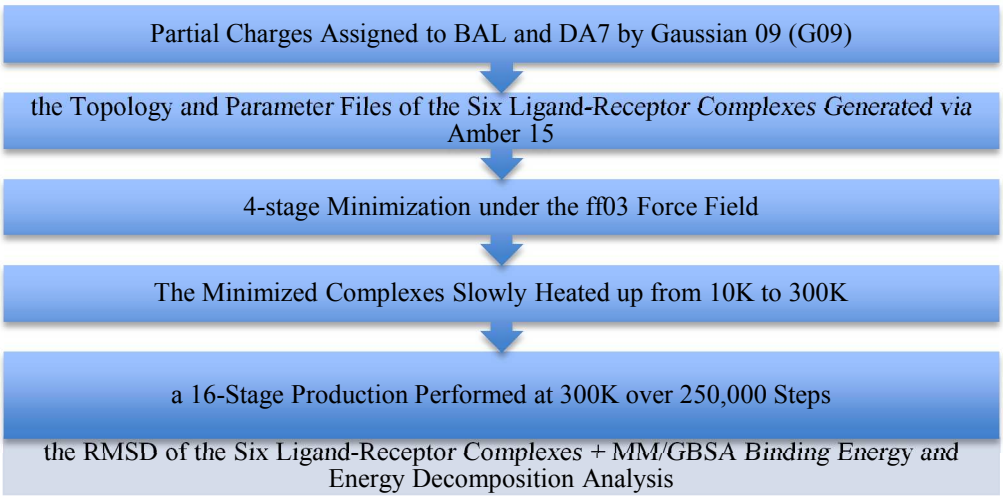
(1)



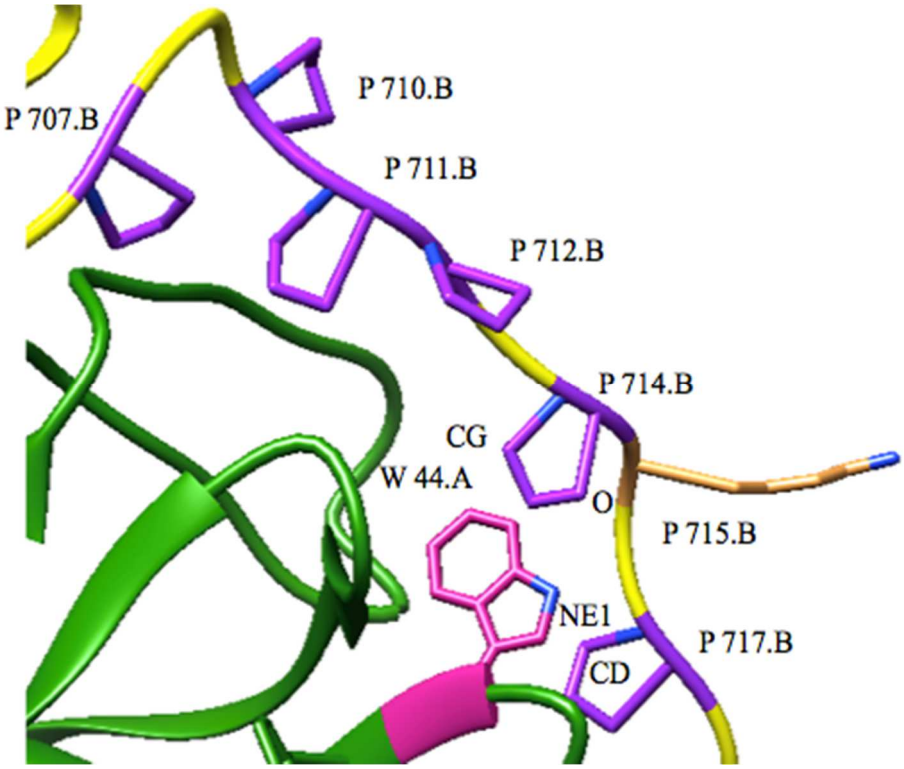
(2)



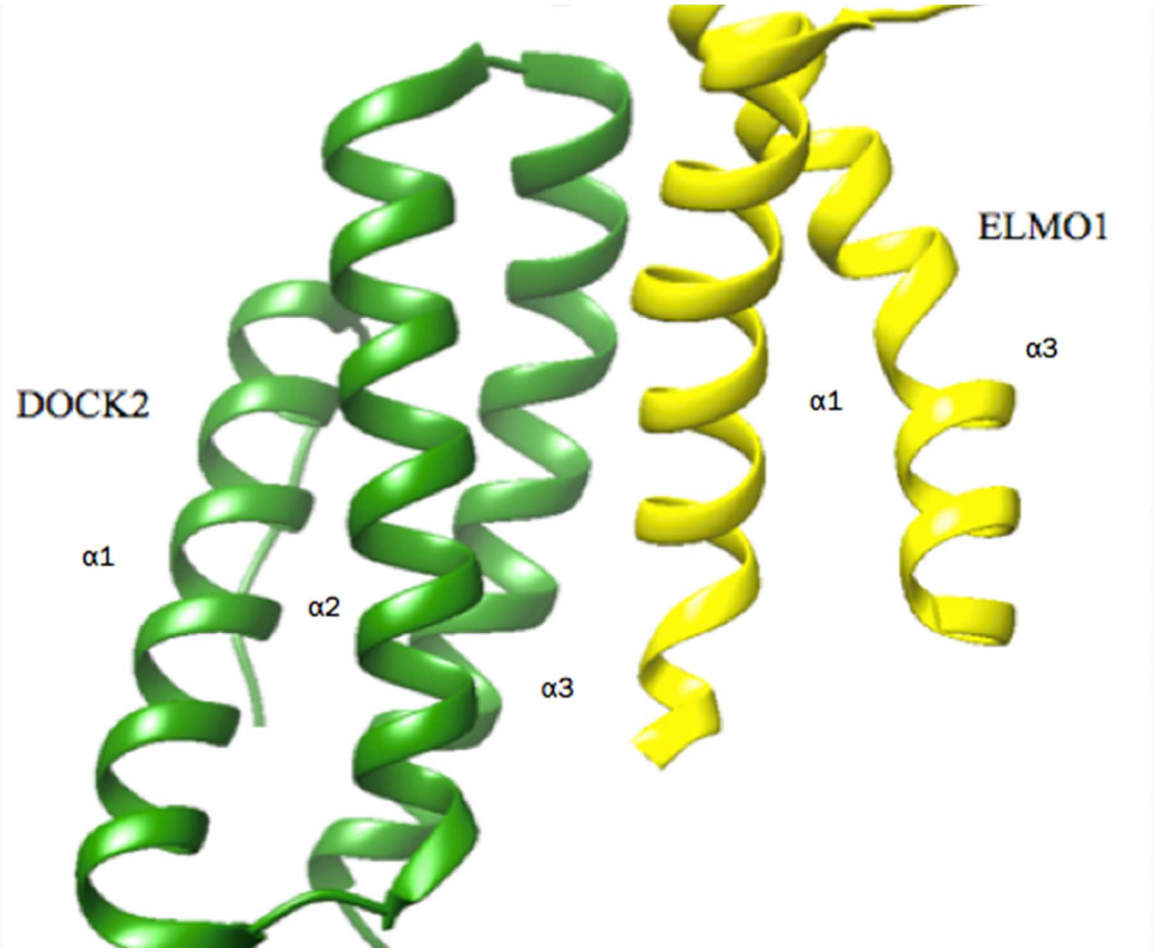
(3)



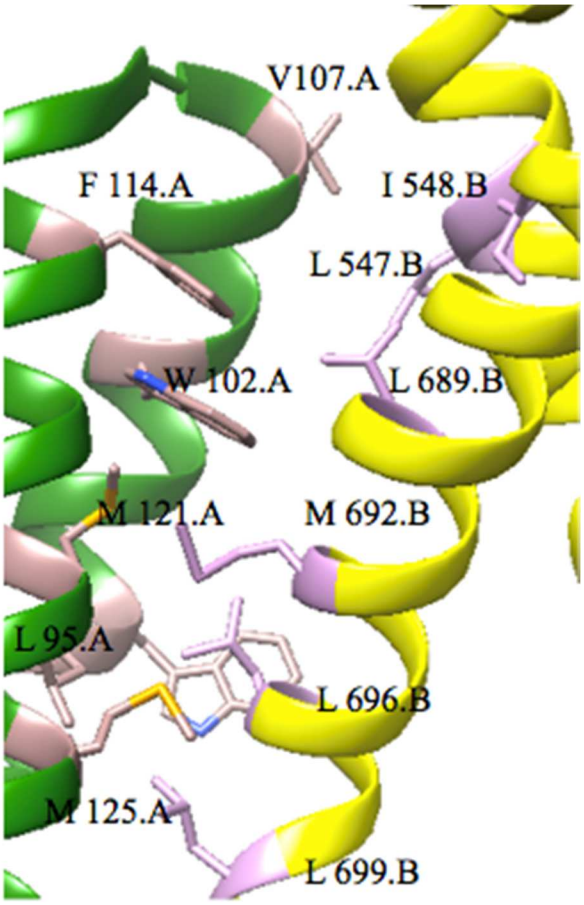
Sup. Figure J. 1-3. shows the flow chart for the following:
Sup. Figure J. 1. Docking process of DAPK2
Sup. Figure J. 2. Ligand similarity analysis of flexible docking results for DAPK2
Sup. Figure J. 3. MD simulation method for the four DAPK-ligand complexes.



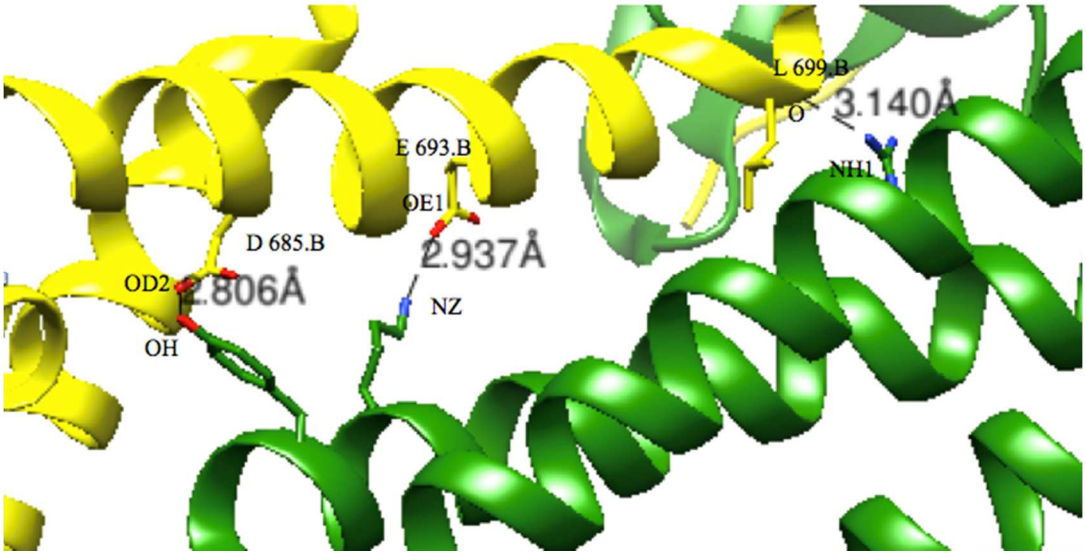
Sup. Figure K. shows the overview of Pxxp-motif of ELMO1 and SH3 domain of DOCK2. Pro 707,710,711,712,714 and 717 are colored in purple. SH3 domain of DOCK2 is colored in green; Trp 44 and Lys715 are colored in violet red and sandy brown respectively. NE1 of Trp44 at the SH3 domain bonds hydrophobically with CG of Pro714 and CD of Pro716 at Pxxp-motif with a distance of 3.571Å, 3.764 Å respectively. In addition, NE1 of Trp44 forms hydrogen bonds with the backbone oxygen located at the Lys715 of ELMO1 with a distance of 2.777 Å.



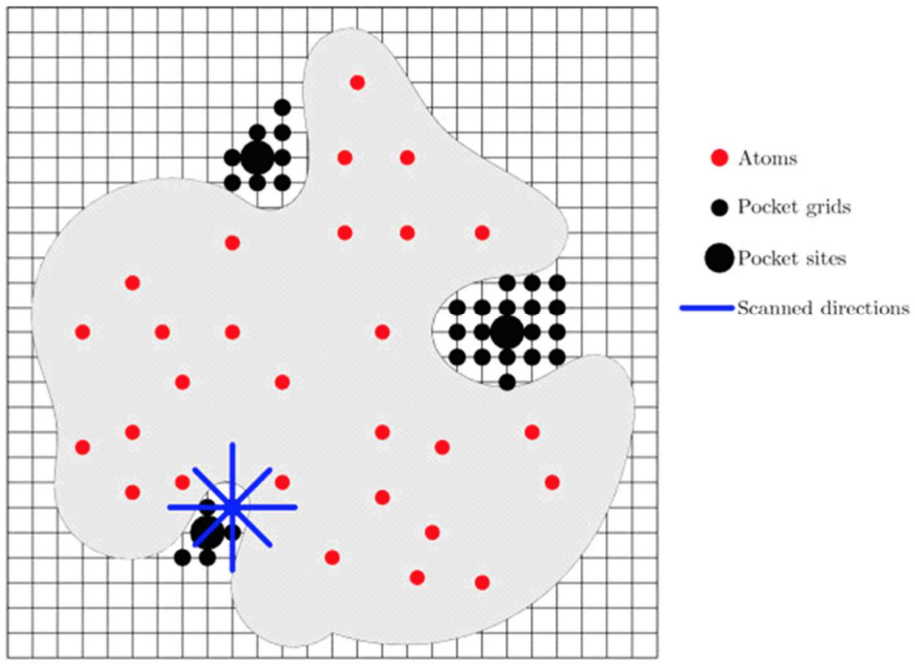
Sup. Figure L. shows the five-bundle helices formed by ELMO1 (yellow) and DOCK2 (green).



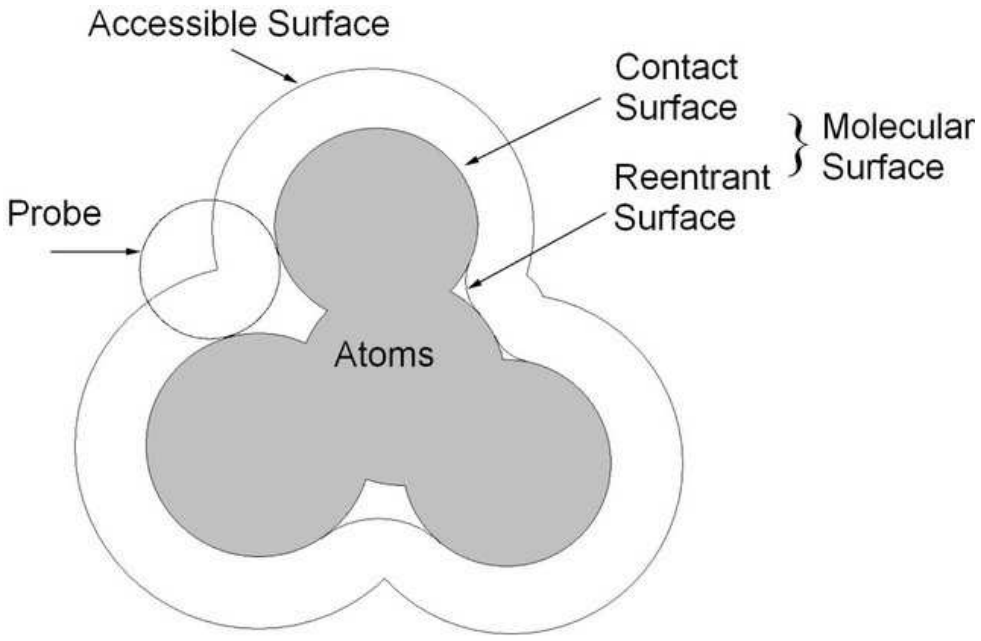
Sup. Figure M. Overview of the hydrophobic interactions formed between in the five-bundle helices. Leu 95, Trp96, Trp 102, Val 107, Phe 114, Met 121 (in rosy brown) and Met 125 at DOCK2 interact with the Ile 544, Leu 547, Ile 548, Leu 681, Leu 689, Met 692 Leu 696, and Leu 699 (in plum) at ELMO1.



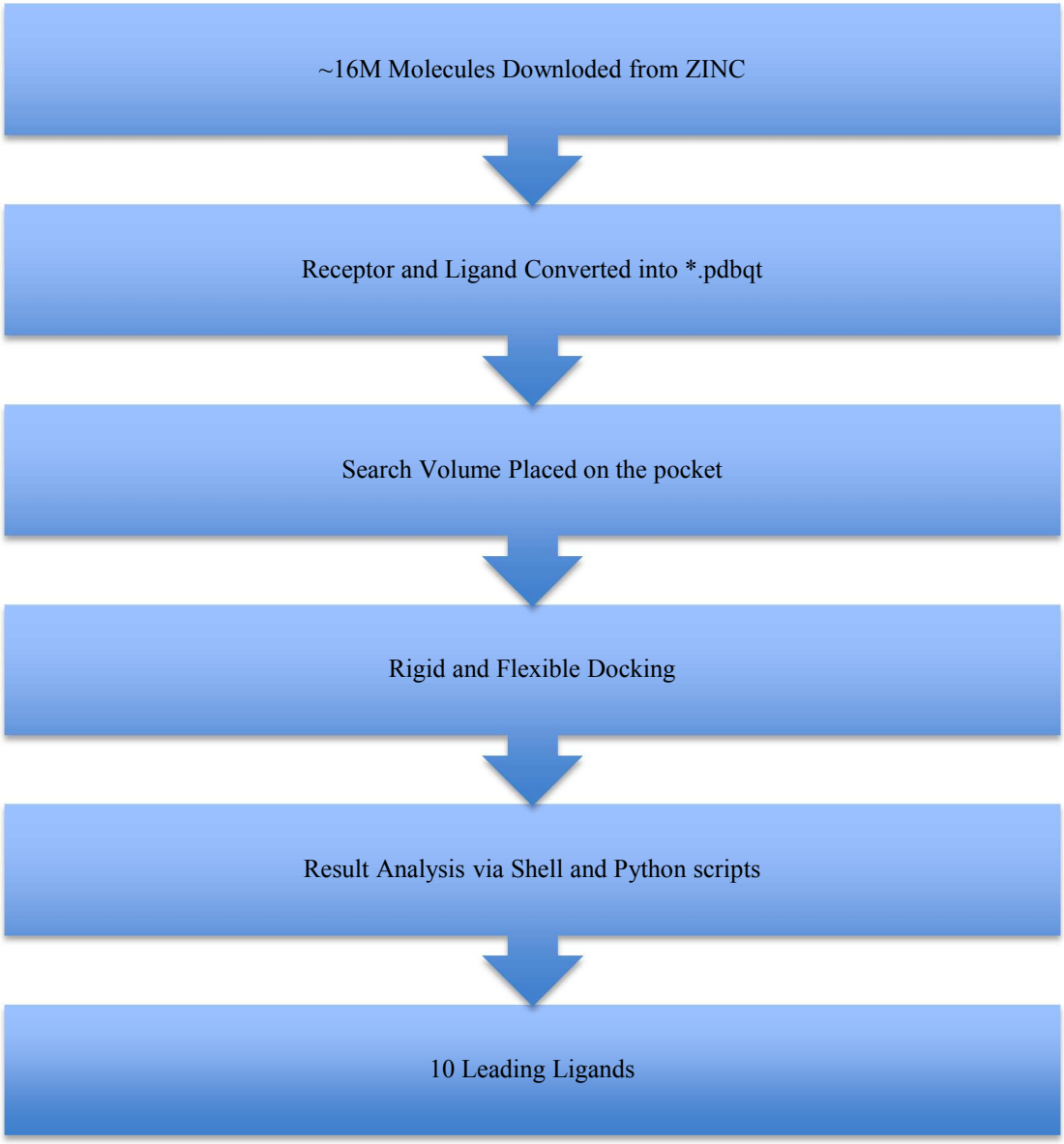
Sup. Figure N. Overview of the hydrophilic interactions formed between in the five-bundle helices. The side chain of Lys103, Tyr 106 and Arg 128 of ELMO1 form hydrogen bond with the side chain of Glu693, Arg 685 and Leu 699 of DOCK2 respectively, with a distance of with a distance of 2.937Å, 2.806Å, and 3.140 Å respectively.



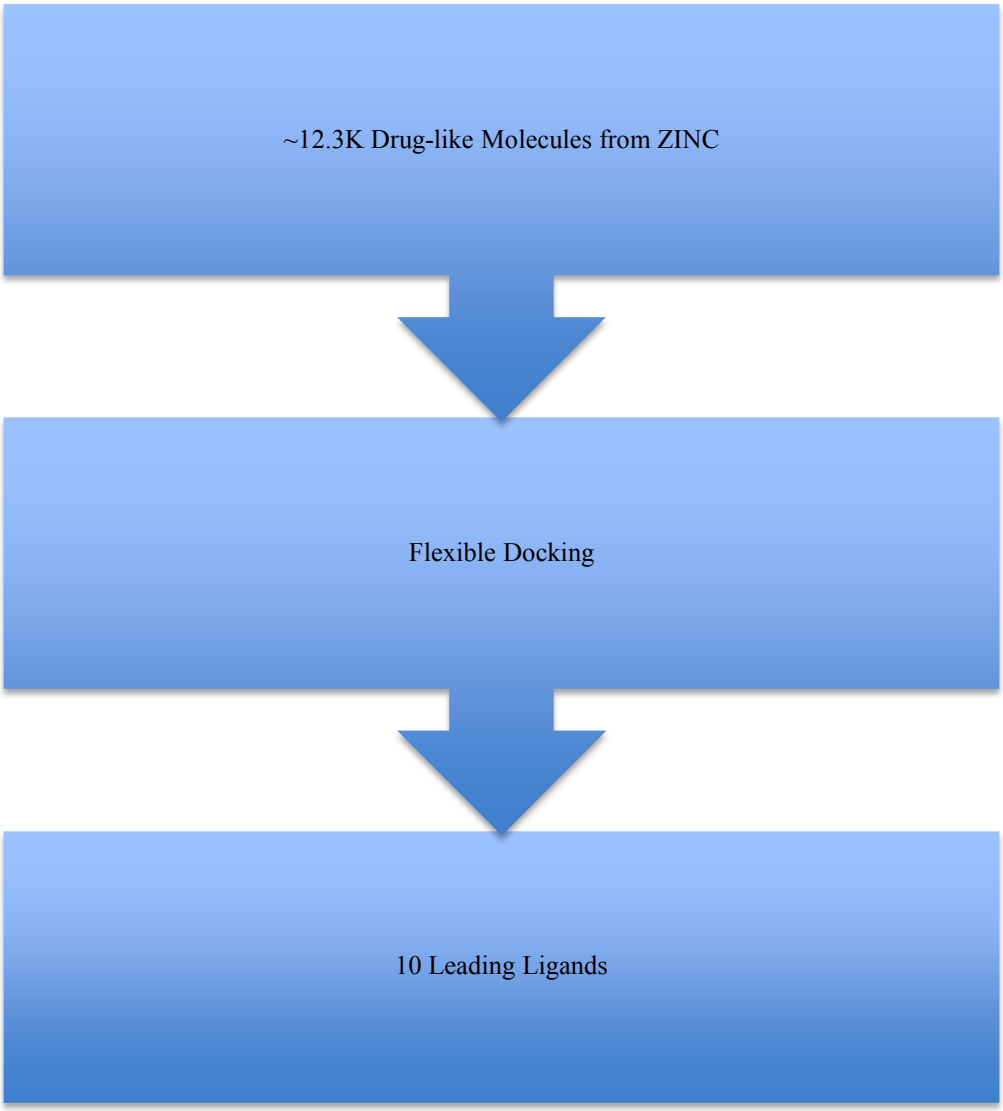
Sup. Figure O. Overview of LIGSITE^{csc} Algorithm adapted from Huang and Michael et.al [74].



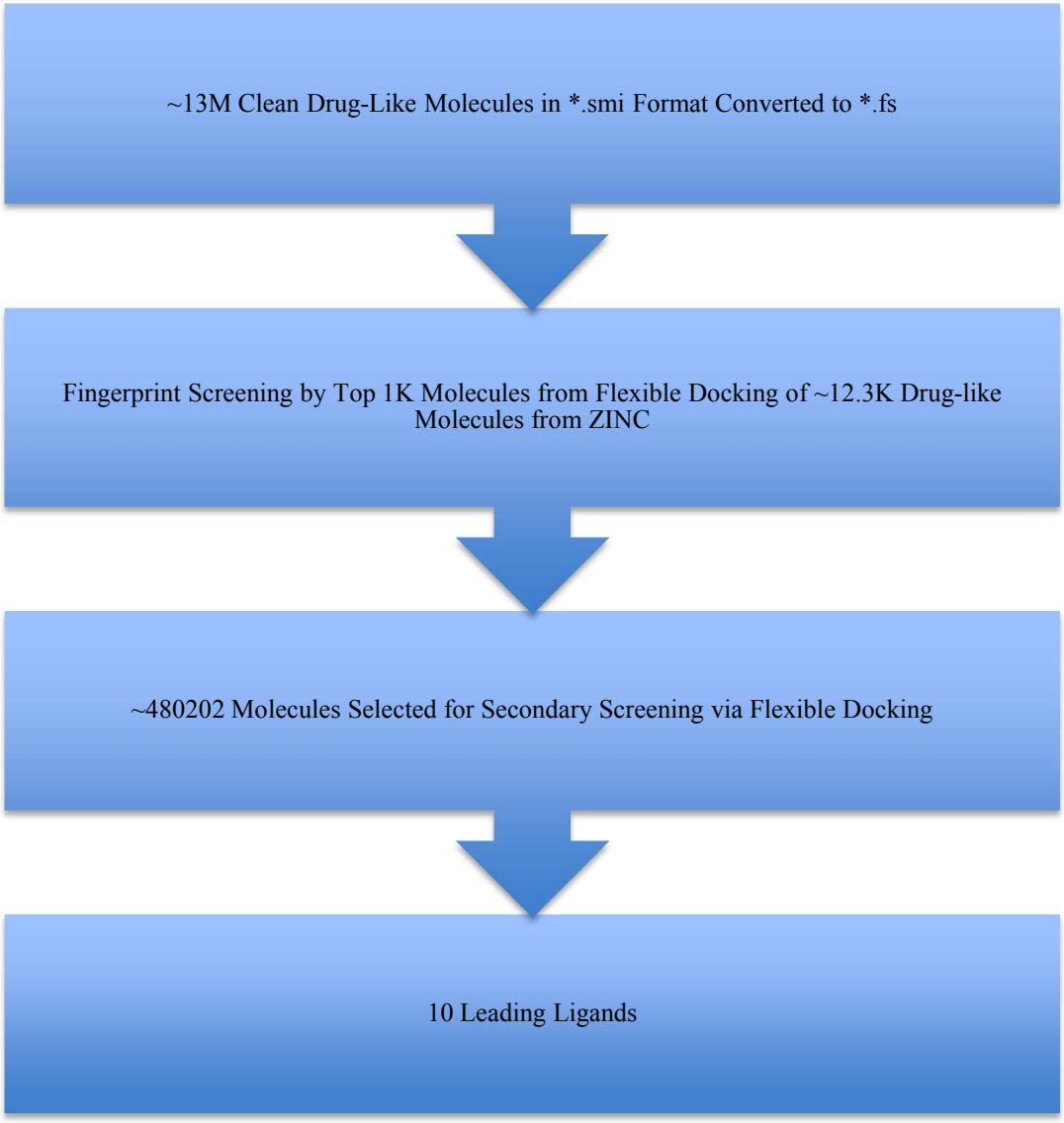
Sup. Figure P. Overview of solvent accessible surface based on Yi.et al. [75]



Sup. Figure Q. shows the flow chart for the docking method used in Step **2.**

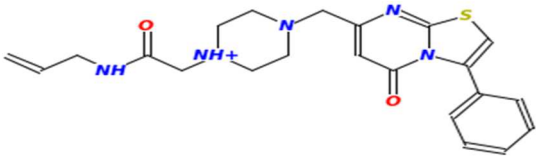


Sup. Figure R. shows the flow chart for the docking method used in Step **3.**

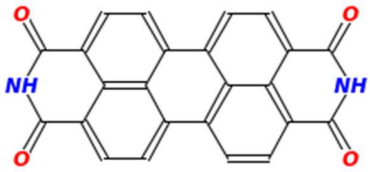


Sup. Figure S. shows the flow chart for the docking method used in Steps **4 and 5.**

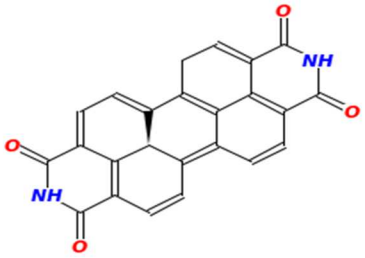
1. ZINC22013692
(-8.8Kcal/mol, 442.478g/mol)



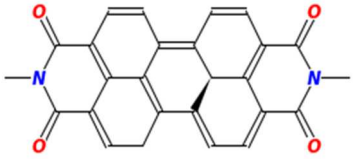
2. ZINC22013688
(-8.6Kcal/mol, 442.478g/mol)



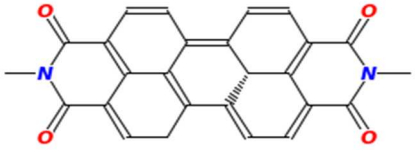
3.ZINC00640220
(-8.6Kcal/mol, 390.354g/mol)



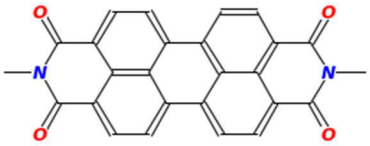
4. ZINC71257150
(-8.5Kcal/mol, 392.37g/mol)



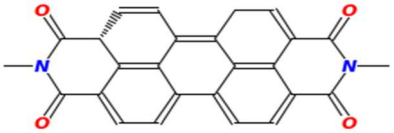
5.ZINC71257153
(-8.4Kcal/mol, 420.424g/mol)



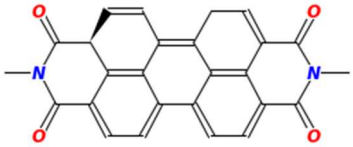
6.ZINC71257152
(-8.4Kcal/mol, 420.424g/mol)



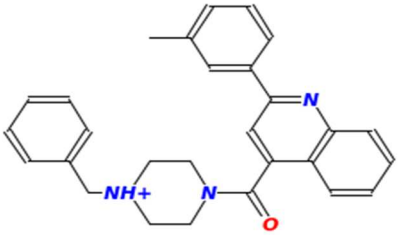
7.ZINC11612116
(-8.3Kcal/mol, 418.408g/mol)



8.ZINC22003706
(-8.3Kcal/mol, 420.424g/mol)

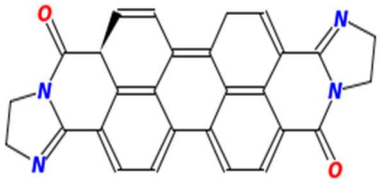


9.ZINC22003711
(-8.2Kcal/mol, 420.424g/mol)

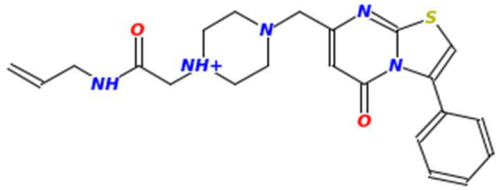


10. ZINC37865607
(-8.1Kcal/mol, 358.421g/mol)

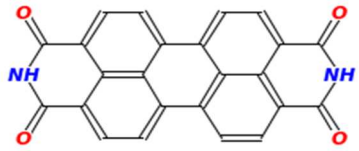
Sup. Figure T. shows the top 10-ligand candidates for ELMO1-DOCK2 complex at SH3 Pocket in rigid docking, with their corresponding ZINC ID (exhaustiveness = 1). These ligands were numbered based on their binding affinity.



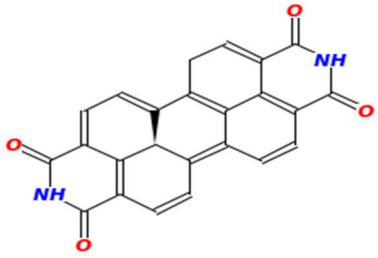
1. ZINC22013692
(-8.8Kcal/mol, 442.478g/mol)



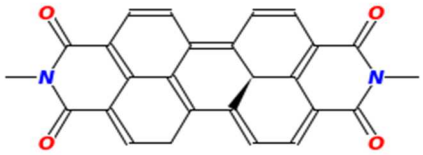
2. ZINC22013688
(-8.6Kcal/mol, 442.478g/mol)



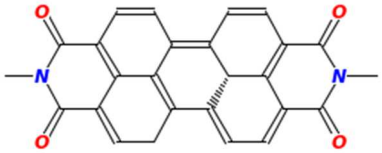
3.ZINC00640220
(-8.6Kcal/mol, 390.354g/mol)



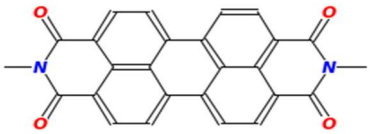
4. ZINC71257150
(-8.5Kcal/mol, 392.37g/mol)



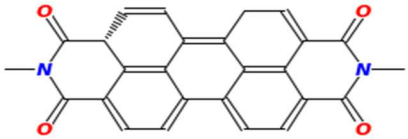
5.ZINC71257153
(-8.4Kcal/mol, 420.424g/mol)



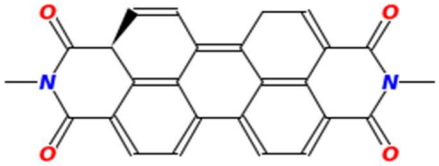
6.ZINC71257152
(-8.4Kcal/mol, 420.424g/mol)



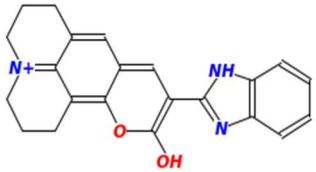
7.ZINC11612116
(-8.3Kcal/mol, 418.408g/mol)



8.ZINC22003706
(-8.3Kcal/mol, 420.424g/mol)

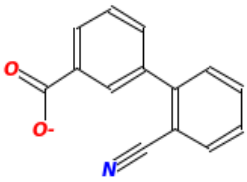


9.ZINC22003711
(-8.2Kcal/mol, 420.424g/mol)

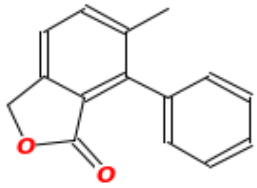


10. ZINC37865607
(-8.1Kcal/mol, 422.552g/mol)

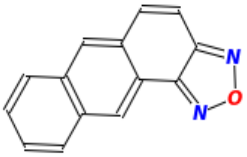
Sup. Figure U. Top 10 Ligand Candidates for ELMO1-DOCK2 complex at SH3 Pocket in Flexible Docking, with their corresponding ZINC ID (exhaustiveness = 9). These ligands were numbered based on their binding affinity. Continue.



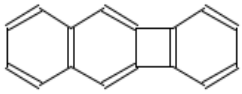
1.ZINC16947599
(-7.0Kcal/mol, 222.223g/mol)



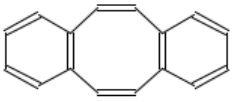
2.ZINC01618286
(-7.0Kcal/mol, 224.259g/mol)



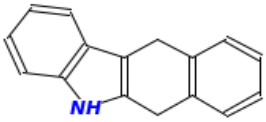
3.ZINC05777952
(-6.9Kcal/mol, 220.231g/mol)



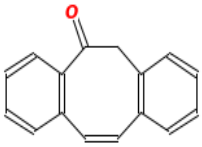
4.ZINC03124423
(-6.8Kcal/mol, 202.256g/mol)



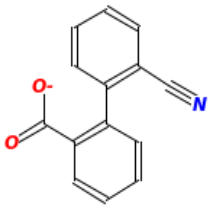
5.ZINC68565001
(-6.8Kcal/mol, 204.272g/mol)



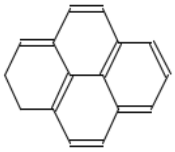
6.ZINC71404348
(-6.8Kcal/mol, 219.287g/mol)



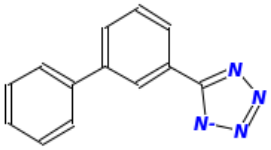
7.ZINC16954355
(-6.7Kcal/mol, 220.271g/mol)



8.ZINC00405763
(-6.7Kcal/mol, 222.223g/mol)

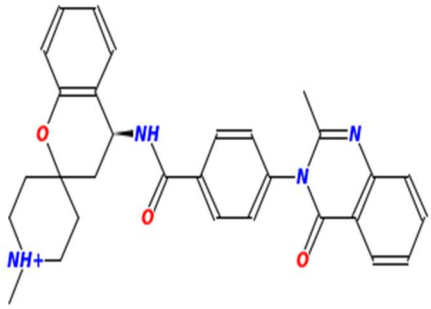


9.ZINC01720161
(-6.7Kcal/mol, 204.272g/mol)

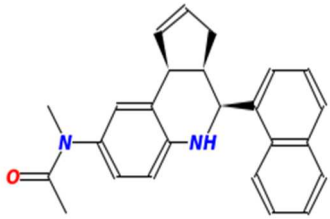


10.ZINC15043023
(-6.6Kcal/mol, 221.243g/mol)

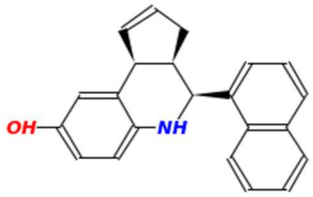
Sup. Figure V. Top 10 Ligand Candidates for ELMO1-DOCK2 at helix pocket based on their binding affinity, with its corresponding ZINC ID, among the ~12.3K Clean Drug-Like Molecules with Tanimoto coefficient = 0.9. These ligands were numbered based on their binding affinity. Since the binding affinity wasn't ideal enough to function, as the inhibitors for ELMO1-DOCK2 complex, a secondary ligand-based screening was conducted.



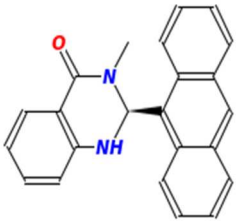
1.ZINC48368120
(-8.3Kcal/mol, 482.668g/mol)



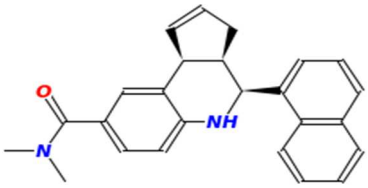
2. ZINC20698566
(-8.2Kcal/mol, 368.48g/mol)



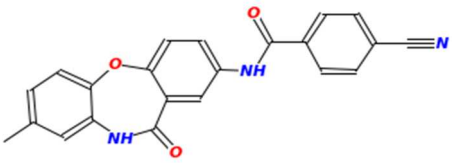
3.ZINC14463638
(-8.1Kcal/mol, 313.4g/mol)



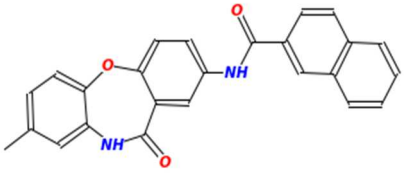
4.ZINC66128373
(-8.1Kcal/mol, 338.41g/mol)



5.ZINC20699276
(-8.1Kcal/mol, 368.48g/mol)

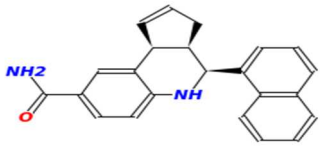


6.ZINC38715363
(-8.0Kcal/mol 369.38g/mol)

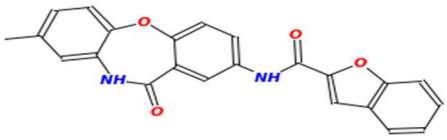


7.ZINC38715402
(-8.0Kcal/mol, 394.43g/mol)

8.ZINC38715378
(-8.0Kcal/mol , 380.35g/mol)



9.ZINC3871538010
(-8.0Kcal/mol, 340.426g/mol)



10.ZINC14463744
(-8.0Kcal/mol, 384.391g/mol)

Sup. Figure W. shows the top ten ligand-candidates for ELMO1-DOCK2 complex at helix pocket during secondary ligand-based screening, in flexible docking (Exhaustiveness = 9). These ligand candidates were ranked solely by the binding affinity, among the ~480,202 Clean-Fragment Like molecules from Step **5**.